

University of London

State Space Models: Univariate  
Representation of a Multivariate Model,  
Partial Interpolation and Periodic  
Convergence

Miltiadis C. Mavrakakis

London School of Economics and Political Science

Thesis submitted for the degree of Doctor of Philosophy

*Department of Statistics, Houghton Street, London WC2A 2AE*

June 2008

UMI Number: U615289

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U615289

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

THESES

F

9052



1198280

*Time, they say, is the answer. But I don't believe them.*

Sly Stone, *Time*

## Abstract

This thesis examines several issues that arise from the state space representation of a multivariate time series model.

Original proofs of the algorithms for obtaining interpolated estimates of the state and observation vectors from the Kalman filter smoother (KFS) output are presented, particularly for the formulae for which rigorous proofs do not appear in the existing literature. The notion of partially interpolated estimates is introduced and algorithms for constructing these estimates are established.

An existing method for constructing a univariate representation (UR) of a multivariate model is developed further, and applied to a wider class of state space models. The computational benefits of filtering and smoothing with the UR, rather than the original multivariate model, are discussed. The UR KFS recursions produce useful quantities that cannot be obtained from the original multivariate model. The mathematical properties of these quantities are examined and the process of reconstructing the original multivariate KFS output is demonstrated.

By reversing the UR process, a time-invariant state space form (SSF) is proposed for models with periodic system matrices. This SSF is used to explore the novel concept of periodic convergence of the KFS. Necessary and sufficient conditions for periodic convergence are asserted and proved.

The techniques developed are then applied to the problem of missing-value estimation in long multivariate temperature series, which can arise due to gaps in the historical records. These missing values are a hindrance to the study of weather risk and pricing of weather derivatives, as well as the development of climate-dependent models. The proposed model-based techniques are compared to existing methods in the field, as well as an original ad hoc approach.

The relative performance of these methods is assessed by their application to data from weather stations in the state of Texas, for daily maximum temperatures from 1950 to 2001.

# Acknowledgements

Firstly, I would like to thank my supervisor, Dr. Jeremy Penzer, whose guidance, insight and patience made this thesis possible. I cannot imagine having a better mentor for my PhD. I also thank my second supervisor, Prof. Qiwei Yao, for his valuable and perceptive advice.

Thanks to my undergraduate Director of Studies, Dr. Mark Spivack, for his faith in me and for encouraging me to pursue research.

Thanks to the whole LSE Statistics Department, particularly to Imelda and Esther, for always helping me sort out my practical problems, however self-caused they may have been.

Thanks to my ST102 and ST202 students for always testing my knowledge of statistics and causing me to reassess my ideas.

I thank my colleagues Yorghos, Panos and Diego for welcoming me to LSE and making my start as smooth and pleasant as possible. Thanks also to Neil, Adrian, Billy and everyone else I shared an office with, for making my stay enjoyable.

Thanks to all my friends in London, and especially Zara, for constantly nudging me in the right direction. I would also like to thank Dan, without whose encouragement I never would have attempted this.

Thanks to my girlfriend, Pam, for her patience and emotional support. Lastly, thanks to my parents, who have given me the freedom to make my own decisions.

*To my parents, Kostas and Maria; and to Pam*

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Technical Introduction</b>	<b>3</b>
2.1	Introduction . . . . .	3
2.2	Autoregressive Models . . . . .	4
2.2.1	AR( $p$ ) . . . . .	4
2.2.2	VAR( $p$ ) . . . . .	5
2.2.3	SUAR( $p$ ) . . . . .	5
2.3	Structural Time Series Models . . . . .	6
2.3.1	Introduction . . . . .	6
2.3.2	Examples of Structural Models . . . . .	6
2.4	State Space Representation . . . . .	8
2.4.1	Definitions . . . . .	8
2.4.2	Autoregressive Models in State Space Form . . . . .	10
2.5	The Kalman Filter . . . . .	13
2.5.1	Introduction . . . . .	13
2.5.2	Linear Estimation . . . . .	13
2.5.3	Derivation . . . . .	17



2.5.4	Cholesky Decomposition . . . . .	22
2.5.5	Model Estimation . . . . .	23
2.6	The Kalman Filter Smoother . . . . .	24
2.6.1	Fixed-Interval Smoothing . . . . .	24
2.6.2	Disturbance Smoothing . . . . .	28
<b>3</b>	<b>Interpolated Estimates</b>	<b>33</b>
3.1	Introduction . . . . .	33
3.2	Interpolation . . . . .	34
3.2.1	Introduction . . . . .	34
3.2.2	Observation Estimate . . . . .	34
3.2.3	State Estimate . . . . .	43
3.3	Partially Interpolated Estimates . . . . .	47
3.3.1	Introduction . . . . .	47
3.3.2	Observation Estimate . . . . .	48
3.3.3	State Estimate . . . . .	55
3.4	Conclusions . . . . .	58
<b>4</b>	<b>The Univariate Representation of a Multivariate Time Series</b>	<b>60</b>
4.1	Introduction . . . . .	60
4.2	Motivation . . . . .	61
4.3	Reformulating the State Space Model . . . . .	62
4.3.1	The Koopman-Durbin Approach . . . . .	62
4.3.2	Non-diagonal Measurement Error Variance . . . . .	63

4.3.3	A More General State Space Model . . . . .	64
4.3.4	Correlation between Measurement Error and Transition error . . . . .	65
4.4	Filtering . . . . .	67
4.4.1	Univariate Filtering . . . . .	67
4.4.2	Filtered State Estimate . . . . .	69
4.4.3	Innovations . . . . .	71
4.4.4	Innovation Variance . . . . .	73
4.5	Smoothing . . . . .	75
4.5.1	Univariate Smoothing . . . . .	75
4.5.2	Deletion Residuals . . . . .	78
4.6	Contemporaneous Observations . . . . .	81
4.6.1	Introduction . . . . .	81
4.6.2	The General Model . . . . .	81
4.6.3	Filtering . . . . .	82
4.7	Conclusions . . . . .	86
<b>5</b>	<b>The Steady-State Filter and Periodic Convergence</b>	<b>88</b>
5.1	Introduction . . . . .	88
5.2	Filter Steady State . . . . .	89
5.2.1	Time-invariant Models . . . . .	89
5.2.2	Definitions . . . . .	91
5.2.3	Conditions for Convergence . . . . .	93
5.2.4	General SSF . . . . .	94

5.2.5	Application . . . . .	97
5.3	Generalisation to Periodic Models . . . . .	99
5.3.1	Introduction . . . . .	99
5.3.2	Periodic Convergence . . . . .	100
5.3.3	Stability . . . . .	106
5.3.4	Detectability and Observability . . . . .	107
5.3.5	Stabilisability . . . . .	110
5.4	Conclusions . . . . .	113
<b>6</b>	<b>Estimation of Missing Temperature Data</b>	<b>114</b>
6.1	Introduction . . . . .	114
6.1.1	Definitions . . . . .	115
6.2	Ad-hoc Approach . . . . .	117
6.2.1	Existing Methods . . . . .	117
6.2.2	The Harmonic Model . . . . .	120
6.2.3	$k$ -Nearest Neighbour Methods . . . . .	131
6.3	Modelling Daily Temperature Series . . . . .	135
6.3.1	Motivation . . . . .	135
6.3.2	The ARFIMA Model . . . . .	135
6.3.3	Estimation of $d$ - Aggregated Variance . . . . .	137
6.3.4	Differenced Variance . . . . .	139
6.3.5	Maximum-Likelihood Methods . . . . .	141
6.3.6	Removing Long Memory . . . . .	142
6.3.7	Model-based Estimation . . . . .	143

6.4	Results and Conclusions . . . . .	145
6.4.1	Station selection . . . . .	145
6.4.2	Ad-hoc Methods . . . . .	146
6.4.3	Model-based Methods . . . . .	148
<b>A</b>	<b>Figures</b>	<b>150</b>
<b>B</b>	<b>Tables</b>	<b>155</b>

# List of Figures

A.1	Periodogram of $\hat{y}_{i,t}$ (Lubbock, TX) . . . . .	150
A.2	Periodogram of $(\zeta_{i,t})$ (Lubbock, TX) . . . . .	151
A.3	Periodogram of $(\zeta_{i,t}^*)^2$ (Lubbock, TX) . . . . .	151
A.4	Periodogram of $z_{i,t}$ (Lubbock, TX) . . . . .	152
A.5	ACF of the deseasonalised series (Lubbock, TX) . . . . .	152
A.6	Aggregated Variance log-log plot (Lubbock, TX) . . . . .	153
A.7	Partial ACF of $\{w_{i,t}\}$ (Lubbock, TX) . . . . .	153
A.8	Map of Texas . . . . .	154

# List of Tables

B.1	Summary statistics (El Paso, TX and Lubbock, TX)	155
B.2	Results from existing estimation methods (El Paso, TX)	155
B.3	Results from existing estimation methods (Lubbock, TX)	156
B.4	Results from Harmonic Departures methods (El Paso, TX)	156
B.5	Results from Harmonic Departures methods (Lubbock, TX)	156
B.6	Results from NN Methods (El Paso, TX)	156
B.7	Results from NN Methods (Lubbock, TX)	157
B.8	Results from model-based methods (El Paso, TX)	157
B.9	Results from model-based methods (Lubbock, TX)	157

# Chapter 1

## Introduction

In recent years, structural models and the state space form (SSF) have had a profound effect on time series analysis. Together, they form a complete framework for the analysis of time series which can be used as an alternative to the traditional ARIMA methodology.

ARIMA modelling, also known as the Box-Jenkins approach, after the work of Box & Jenkins (1970), relies on differencing a time series to eliminate trend, seasonality and other such time-varying behaviour. The aim is to reduce the data to a stationary series, that is, one where the moments (mean, covariance, etc.) are time-invariant.

Box-Jenkins models can be difficult to identify, since the model choice depends entirely on the data, not the structure of the system. The model parameters relate to abstract quantities and rarely have an intuitive interpretation. In addition, the models can be difficult to estimate, particularly if there are missing observations in the time series. Model fitting is heavily reliant on the sample autocorrelation function (ACF) and partial autocorrelation function (PACF), both of which can be very noisy. Perhaps most importantly, the process of differencing to achieve stationarity tends to eliminate the salient features of the time series.

The basis of state space modelling can be found in the work of Kalman (1960). Initial developments took place in the field of engineering, particularly the branch of control theory. The pioneering work of Harvey (1989) introduced structural models and applied the concept of state space modelling within the framework of time series analysis. With

the structural approach, the different components (trend, seasonality, explanatory variables, interventions) are modelled separately and then combined to form the complete model.

Structural models provide explicit information about the individual components, such as the seasonal pattern, which are very often of interest in themselves. The process of model selection is more natural, because the components reflect the real-life structure of the time series. Furthermore, the transparency of the structural approach makes it easy to check the components individually and ensure that their behaviour is realistic. The model parameters relate to interpretable quantities and can be allowed to evolve over time, in marked contrast to the rigidity of the ARIMA approach. Estimation, forecasting and interpolation are all straightforward once a structural time series model is in state space form, by use of the Kalman filter recursions.



## Chapter 2

# Technical Introduction

### 2.1 Introduction

In this chapter we provide the reader with the theoretical background of the thesis. We begin by examining a subclass of Box-Jenkins models, that of autoregressive models, and describing the process by which these are estimated. We are particularly interested in the different multivariate extensions of the the univariate model.

In subsequent sections, we introduce structural time series analysis and give examples of different structural models. We define the state space representation of a time series, focusing on the more general state space form (SSF) introduced by de Jong (1991), and demonstrate that structural and Box-Jenkins models can be expressed in SSF.

Later in the chapter, we consider the Kalman filter (KF) and its importance in model estimation, filtering and forecasting. We prove the key linear estimation results which then allow us to derive the KF recursions. We identify the parallels between the KF and the Cholesky decomposition, which serve to explain the computational efficiency of the KF.

We conclude by discussing the problem of smoothing, that is, obtaining parameter estimates conditional on the entire sample. We derive the Kalman filter smoother (KFS) recursions, which, by augmenting the KF, allow us to obtain these estimates in a straightforward manner.

## 2.2 Autoregressive Models

Autoregressive models are ones where each observation is considered to be equal to a weighted sum of past observations plus an error term. They have several advantages: specifying a parsimonious model is simple, and we can often use standard techniques from linear regression theory to fit it to the data. Obtaining forecasts is easy and, in many cases, the models produce reasonably good short-term predictions. Furthermore, generalisation to multivariate models (such as VAR and SUAR, discussed later in this section) is straightforward.

### 2.2.1 AR( $p$ )

A  $p^{th}$ -order autoregression, denoted AR( $p$ ), has model equation

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} \dots + \phi_p y_{t-p} + \varepsilon_t \quad (2.1)$$

where  $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$ . Using the backshift operator  $B$ , (2.1) can be expressed as

$$\Phi(B)y_t = \varepsilon_t \quad (2.2)$$

where  $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  is the *autoregressive polynomial*. If the roots of  $\Phi(z)$  lie outside the unit circle, then the series  $\{y_t\}$  is stationary. The parameters  $\phi_1, \dots, \phi_p$  are estimated by ordinary least squares (OLS), regressing  $y_t$  on its  $p$  first lags. The AR( $p$ ) model does not satisfy the basic assumptions of least squares regression (for example, the predictor variables are stochastic and depend on past errors), so the resulting parameter estimates are biased. However, the estimates are consistent and yield asymptotically valid tests of hypotheses about the autoregressive parameters (e.g. Hamilton, 1994).

### 2.2.2 VAR( $p$ )

A  $p^{th}$ -order vector autoregression, denoted VAR( $p$ ), has model equation

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} \dots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \quad (2.3)$$

where  $\mathbf{y}_t$  is the vector of observations at time  $t$ , the autoregressive coefficients  $\Phi_1, \dots, \Phi_p$  are square matrices, and the error vector  $\{\boldsymbol{\varepsilon}_t\} \sim \text{WN}(\mathbf{0}, \Sigma_\varepsilon)$  has the same length as  $\mathbf{y}_t$ . As before, the autoregressive parameters are estimated by OLS, regressing  $\mathbf{y}_t$  on its  $p$  first lags.

### 2.2.3 SUAR( $p$ )

A system of  $p^{th}$ -order seemingly unrelated autoregressions, denoted by SUAR( $p$ ), consists of the autoregressive processes

$$y_{i,t} = \phi_{i,1} y_{i,t-1} + \phi_{i,2} y_{i,t-2} \dots + \phi_{i,p} y_{i,t-p} + \varepsilon_{i,t} \quad (2.4)$$

for  $i = 1, \dots, k$ , where  $\{\boldsymbol{\varepsilon}_t\} = \{(\varepsilon_{1,t}, \dots, \varepsilon_{k,t})'\} \sim \text{WN}(\mathbf{0}, \Sigma_\varepsilon)$ . The matrix  $\Sigma_\varepsilon$  is assumed to be non-diagonal, thus contemporaneous errors in the different processes are correlated. Note that a SUAR( $p$ ) model is equivalent to a VAR( $p$ ) model where the coefficient matrices  $\Phi_1, \dots, \Phi_p$  are diagonal.

Let  $\phi_i = (\phi_{i,1} \dots \phi_{i,p})'$  and  $\mathbf{x}_{i,t} = (y_{i,t-1} \dots y_{i,t-p})$ , so we can express equation (2.4) as  $y_{i,t} = \mathbf{x}_{i,t} \phi_i + \varepsilon_{i,t}$ . We also define  $\mathbf{y}_t = (y_{1,t} \dots y_{k,t})'$  and  $\mathbf{X}_t = (\mathbf{x}'_{1,t} \dots \mathbf{x}'_{k,t})'$ . SUAR models are commonly estimated using a method introduced by Zellner (1962):

1. Obtain an initial (OLS) set of estimates  $\hat{\phi}_i^{(0)}$  by regressing  $y_{i,t}$  on its first  $p$  lags,  $\mathbf{x}_{i,t}$ .
2. Construct the residuals  $\hat{\varepsilon}_{i,t}^{(l)} = y_{i,t} - \mathbf{x}_{i,t} \hat{\phi}_i^{(l)}$  and calculate the sample variance  $\hat{\Sigma}^{(l)}$  of the vector of residuals  $\hat{\boldsymbol{\varepsilon}}_t^{(l)} = (\hat{\varepsilon}_{1,t}^{(l)} \dots \hat{\varepsilon}_{k,t}^{(l)})'$ .

3. Compute the Cholesky decomposition  $\hat{\Sigma}^{(l)} = \mathbf{L}^{(l)}(\mathbf{L}^{(l)})'$ .
4. Obtain an updated set of estimates  $\hat{\phi}_i^{(l+1)}$  by OLS regression of  $y_{i,t}^{(l)} = \{(\mathbf{L}^{(l)})^{-1} \mathbf{y}_t\}_i$  on  $x_{i,t}^{(l)} = \{(\mathbf{L}^{(l)})^{-1} \mathbf{X}_t\}_i$

where  $\{\dots\}_i$  denotes the  $i^{th}$  row of a matrix. Steps 2 – 4 are repeated until the algorithm converges. The initial OLS parameter estimates are biased, but all further estimates  $\hat{\phi}_i^{(1)}, \hat{\phi}_i^{(2)}, \dots$  are unbiased.

## 2.3 Structural Time Series Models

### 2.3.1 Introduction

A structural time series model is set up in terms of directly interpretable components, which are then modelled as stochastic processes. Each observation  $y_t$  is expressed as a sum of unobserved components, such as the current level, seasonal effect and noise. It is common to assume that the unobserved components are uncorrelated, but this is not strictly necessary; in fact, the relationship between these components can be modelled explicitly.

### 2.3.2 Examples of Structural Models

We will now consider some basic structural models. These are the building blocks of structural time series analysis and can be used to construct more complex models.

#### Local Level Model (LLM)

The simplest structural model is one where the mean of a time series follows a random walk. This is known as the Local Level Model or a "random walk plus noise":

$$\begin{aligned}
y_t &= \mu_t + \varepsilon_t \\
\mu_{t+1} &= \mu_t + \eta_t
\end{aligned} \tag{2.5}$$

where  $\mu_t$  is the level and the error series  $\{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2)$ ,  $\{\eta_t\} \sim \text{WN}(0, \sigma_\eta^2)$  are assumed to be mutually uncorrelated at all lags.

### Local Linear Trend (LLT)

If we include a slope term which also follows a random walk, we obtain the Local Linear Trend model:

$$\begin{aligned}
y_t &= \mu_t + \varepsilon_t \\
\mu_{t+1} &= \mu_t + \beta_t + \eta_t \\
\beta_{t+1} &= \beta_t + \zeta_t
\end{aligned} \tag{2.6}$$

where  $\beta_t$  is the slope and  $\mu_t$  is the trend. The error series  $\{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2)$ ,  $\{\eta_t\} \sim \text{WN}(0, \sigma_\eta^2)$ ,  $\{\zeta_t\} \sim \text{WN}(0, \sigma_\zeta^2)$  are mutually uncorrelated at all lags.

### Dummy Seasonal Model

A straightforward way to model seasonal time series is to introduce additive seasonal effects at each time period. Let  $\gamma_t$  be the seasonal effect at time  $t$ . We could assume that the seasonal effects sum to zero over a whole period, thus:

$$\gamma_t = - \sum_{j=1}^{s-1} \gamma_{t-j} \tag{2.7}$$

If we include this seasonal term in the LLM we obtain

$$\begin{aligned}
y_t &= \mu_t + \varepsilon_t \\
\mu_{t+1} &= \mu_t + \gamma_t + \eta_t \\
\gamma_{t+1} &= - \sum_{j=1}^{s-1} \gamma_{t+1-j}
\end{aligned} \tag{2.8}$$

The error terms  $\{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2)$ ,  $\{\eta_t\} \sim \text{WN}(0, \sigma_\eta^2)$  are mutually uncorrelated at all lags. It is often useful to allow the seasonal terms to evolve stochastically. We can accomplish this by introducing another error term  $\{\omega_t\} \sim \text{WN}(0, \sigma_\omega^2)$  and setting

$$\gamma_t = - \sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t \tag{2.9}$$

In this case, the sum of seasonal effects over a whole period is equal to this error term.

## 2.4 State Space Representation

### 2.4.1 Definitions

Structural time series models, such as those described in the previous section, are estimated by putting them in state space form. The fundamental concept of the state space approach is that the observation  $\mathbf{y}_t$  can be expressed as a linear function of an unobserved vector  $\boldsymbol{\alpha}_t$  (known as the *state* vector) and an error term  $\boldsymbol{\varepsilon}_t$ . The state vectors follow a first-order autoregression, which has the Markov property. Thus, the SSF can be thought of as a Hidden Markov Model.

We will use the SSF defined by de Jong (1991):

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{G}_t \boldsymbol{\varepsilon}_t \quad (2.10)$$

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t \quad (2.11)$$

for  $t = 1, 2, \dots$ . The dimensions of  $\mathbf{y}_t$ ,  $\boldsymbol{\alpha}_t$  and  $\boldsymbol{\varepsilon}_t$  are  $p \times 1$ ,  $q \times 1$  and  $r \times 1$ , respectively. The errors  $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots$  form a white noise process with unit variance, denoted by  $\{\boldsymbol{\varepsilon}_t\} \sim \text{WN}(0, \mathbf{I}_r)$ . The system matrices  $\mathbf{Z}_t, \mathbf{G}_t, \mathbf{T}_t, \mathbf{H}_t$  are deterministic quantities and depend on unknown parameters, which are estimated by maximum likelihood.

Equation (2.10) is known as the *measurement* equation, and describes the process by which the observation vector is calculated from the corresponding state vector. Equation (2.11) is the *transition* equation, which describes the process by which the state vector evolves over time. We will refer to  $\mathbf{G}_t \boldsymbol{\varepsilon}_t$  as the *measurement error*, and  $\mathbf{H}_t \boldsymbol{\varepsilon}_t$  as the *transition error*. A necessary and sufficient condition for contemporaneous measurement and transition errors to be uncorrelated is  $\mathbf{G}_t \mathbf{H}_t' = 0$ .

Relaxing the restriction on the variance of  $\boldsymbol{\varepsilon}_t$  is not necessary. If for a particular model it is more natural to define an error term with non-unit variance, say  $\{\boldsymbol{\varepsilon}_t\} \sim \text{WN}(0, \mathbf{V}_t)$  then we can set

$$\tilde{\boldsymbol{\varepsilon}}_t = \mathbf{D}_t^{-1} \boldsymbol{\varepsilon}_t, \quad \tilde{\mathbf{G}}_t = \mathbf{G}_t \mathbf{D}_t, \quad \tilde{\mathbf{H}}_t = \mathbf{H}_t \mathbf{D}_t \quad (2.12)$$

where  $\mathbf{D}_t$  is the Cholesky decomposition of  $\mathbf{V}_t$ , that is,  $\mathbf{V}_t = \mathbf{D}_t \mathbf{D}_t'$ . The new error series  $\{\tilde{\boldsymbol{\varepsilon}}_t\}$  is a white noise process with unit variance, as required, and the model is unchanged.

We allow the system matrices to evolve over time; this allows the SSF to represent a wide variety of models. However, many important time series models can be expressed in SSF with time-invariant system matrices. For example, consider the LLT model defined in (2.6). This can be put in SSF by setting

$$\mathbf{Z}_t = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad \mathbf{G}_t = \begin{pmatrix} \sigma_\epsilon & 0 & 0 \end{pmatrix}, \quad \{\epsilon_t\} \sim \text{WN}(0, \mathbf{I}_3)$$

$$\boldsymbol{\alpha}_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}, \quad \mathbf{T}_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{H}_t = \begin{pmatrix} 0 & \sigma_\eta & 0 \\ 0 & 0 & \sigma_\zeta \end{pmatrix} \quad (2.13)$$

In the following subsection we will demonstrate how the autoregressive models described in this chapter can also be cast in SSF.

### 2.4.2 Autoregressive Models in State Space Form

#### AR( $p$ )

Consider the AR( $p$ ) model as a special case of a moving-average autoregressive process, denoted ARMA( $p, q$ ), which includes both an autoregressive and a moving-average component:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \dots + \theta_q \epsilon_{t-q} \quad (2.14)$$

or, in lag operator form

$$\Phi(B)y_t = \Theta(B)\epsilon_t \quad (2.15)$$

where  $\{\epsilon_t\} \sim \text{WN}(0, \sigma^2)$ ,  $\Phi(B)$  is the autoregressive polynomial and  $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  is the *moving-average polynomial*. Similarly to the AR( $p$ ) model, the series  $\{y_t\}$  is stationary if the roots of  $\Phi(z)$  lie outside the unit circle.

We want to express this model in the form



$$\begin{aligned}
y_t &= \mathbf{Z}\boldsymbol{\alpha}_t + \mathbf{G}\boldsymbol{\varepsilon}_t \\
\boldsymbol{\alpha}_{t+1} &= \mathbf{T}\boldsymbol{\alpha}_t + \mathbf{H}\boldsymbol{\varepsilon}_t
\end{aligned} \tag{2.16}$$

The most common state space representations in time series literature (e.g. Brockwell & Davis, 1987; Harvey, 1993; Box *et al.*, 1994; Hamilton, 1994) involve a state vector  $\boldsymbol{\alpha}_t$  of dimension  $m = \max(p, q + 1)$ . One such representation (Box *et al.*, 1994) is

$$\mathbf{Z} = (1, 0, \dots, 0), \quad \mathbf{T} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \phi_m & \phi_{m-1} & \cdots & \cdots & \phi_1 \end{pmatrix}$$

$$\mathbf{G} = 0, \quad \mathbf{H} = (1, \psi_1, \dots, \psi_{m-1})' \tag{2.17}$$

where the  $\{\psi_j\}$  are the leading coefficients in the polynomial expansion of  $\Theta(B)/\Psi(B)$ , and are functions of the hyperparameters. De Jong & Penzer (2004) argue in favour of a  $m = \max(p, q)$  representation originally proposed by Pearlman (1980). This has system matrices

$$\mathbf{Z} = (1, 0, \dots, 0), \quad \mathbf{T} = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \phi_{m-1} & 0 & \cdots & 0 & 1 \\ \phi_m & 0 & \cdots & \cdots & 0 \end{pmatrix}$$

$$\mathbf{G} = 1, \quad \mathbf{H} = (\theta_1 + \phi_1, \dots, \theta_m + \phi_m)' \tag{2.18}$$

Note that in this representation,  $\mathbf{GH}' \neq 0$ , thus measurement and transition errors are correlated. In the special case of an AR( $p$ ) model, the

moving average coefficients are all equal to zero and we set

$$\mathbf{H} = (\phi_1, \dots, \phi_p)' \quad (2.19)$$

The other system matrices are the same as in (2.18).

### VAR( $p$ )

The SSF for the VAR( $p$ ) model is very similar to (2.18); the system matrices can be expressed in block form as

$$\mathbf{Z} = [\mathbf{I}, \mathbf{0}, \dots, \mathbf{0}], \quad \mathbf{T} = \begin{bmatrix} \Phi_1 & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \Phi_2 & \mathbf{0} & \mathbf{I} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \Phi_{p-1} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{I} \\ \Phi_p & \mathbf{0} & \dots & \dots & \mathbf{0} \end{bmatrix}$$

$$\mathbf{G} = \mathbf{I}, \quad \mathbf{H} = [\Phi_1', \dots, \Phi_p']' \quad (2.20)$$

where  $\mathbf{I}$  is the identity matrix (de Jong & Penzer, 2004). If the observation vector  $\mathbf{y}_t$  consists of  $k$  components, then the state vector has length  $kp$ .

### SUAR( $p$ )

Assume we observe  $k$  contemporaneous time series  $\{y_{1,t}\}, \{y_{2,t}\}, \dots, \{y_{k,t}\}$ , for  $t = 1, \dots, n$  and we want to model these as a system of seemingly-unrelated autoregressions. We first define the observation vector  $\mathbf{y}_t = (y_{1,t}, \dots, y_{k,t})'$  and the matrices of parameters  $\Phi_j = \text{diag}(\phi_{1,j}, \dots, \phi_{k,j})$ ,  $j = 1, \dots, p$ , where  $\phi_{i,1}, \dots, \phi_{i,p}$  are the autoregressive parameters corresponding to  $\{y_{i,t}\}$ . The system matrices for the VAR( $p$ ) model in (2.20) then define a suitable state space representation for the SUAR( $p$ ) model.

## 2.5 The Kalman Filter

### 2.5.1 Introduction

In its basic form, the Kalman filter produces a linear estimate of the current state vector and its associated mean squared error (MSE), conditional on all data currently available. However, given the simple Markovian structure of the state recursion, it is trivial to extend this filtered estimate into a forecast of future states. Additionally, in the Gaussian case, the likelihood function can be calculated using the KF output, in a way that results in considerable computational savings.

The Kalman filter smoother (KFS), an additional recursive algorithm closely related to the KF, can be used to estimate past states and error terms conditional on the data currently available. In Chapter 4 we demonstrate that the KFS output is essential for estimating missing observations in a data set (interpolation).

### 2.5.2 Linear Estimation

The following lemmas from linear estimation theory provide the basis for the treatment of the Kalman filter and smoother:

**Lemma 2.5.1.** *Let  $\mathbf{x}, \mathbf{y}$  be random vectors. The minimum mean-square linear estimator (MMSLE) of  $\mathbf{x}$  given  $\mathbf{y}$  is*

$$\mathbf{L}(\mathbf{x}|\mathbf{y}) = \mathbf{E}(\mathbf{x}) + \text{Cov}(\mathbf{x}, \mathbf{y})\text{Var}(\mathbf{y})^{-1}[\mathbf{y} - \mathbf{E}(\mathbf{y})] \quad (2.21)$$

*and its MSE as an estimator of  $\mathbf{x}$  is*

$$\text{MSE}_{\mathbf{x}}[\mathbf{L}(\mathbf{x}|\mathbf{y})] = \text{Var}(\mathbf{x}) - \text{Cov}(\mathbf{x}, \mathbf{y})\text{Var}(\mathbf{y})^{-1}\text{Cov}(\mathbf{y}, \mathbf{x}) \quad (2.22)$$

*Proof.* A proof can be found in Duncan & Horn (1972).

□

If  $\mathbf{x}, \mathbf{y}$  are uncorrelated, then  $L(\mathbf{x}|\mathbf{y}) = E(\mathbf{x})$ . Furthermore, it is trivial to verify that, for random vectors  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}$  and any constant matrix  $\mathbf{C}$ , the MMSLE operator  $L(\cdot|\mathbf{y})$  has the following properties:

$$L(\mathbf{C}\mathbf{x}_1|\mathbf{y}) = \mathbf{C}L(\mathbf{x}_1|\mathbf{y}) \quad (2.23)$$

$$L(\mathbf{x}_1 + \mathbf{x}_2|\mathbf{y}) = L(\mathbf{x}_1|\mathbf{y}) + L(\mathbf{x}_2|\mathbf{y}) \quad (2.24)$$

Thus  $L(\cdot|\mathbf{y})$  is a linear operator. Furthermore, the MMSLE  $L(\mathbf{x}|\mathbf{y})$  is an unbiased estimator of  $\mathbf{x}$ , in the sense that

$$\begin{aligned} E[L(\mathbf{x}|\mathbf{y})] &= E(\mathbf{x}) + \text{Cov}(\mathbf{x}, \mathbf{y})\text{Var}(\mathbf{y})^{-1}[E(\mathbf{y}) - E(\mathbf{y})] \\ &= E(\mathbf{x}) \end{aligned} \quad (2.25)$$

This is analogous to the law of iterated expectations. Because of the similarities between the properties of the MMSLE and the conditional expectation  $E(\mathbf{x}|\mathbf{y})$ , we will also refer to  $L(\mathbf{x}|\mathbf{y})$  as the *linear expectation* of  $\mathbf{x}$  given  $\mathbf{y}$ .

**Corollary 2.5.2.** *The linear estimation error  $\mathbf{x} - L(\mathbf{x}|\mathbf{y})$  is uncorrelated with  $\mathbf{y}$ .*

*Proof.* Directly from formula (2.5.1), we have:

$$\begin{aligned} \text{Cov}[\mathbf{y}, \mathbf{x} - L(\mathbf{x}|\mathbf{y})] &= \text{Cov}[\mathbf{y}, \mathbf{x} - E(\mathbf{x}) - \text{Cov}(\mathbf{x}, \mathbf{y})\text{Var}(\mathbf{y})^{-1}\{\mathbf{y} - E(\mathbf{y})\}] \\ &= \text{Cov}(\mathbf{y}, \mathbf{x}) - \text{Cov}\{\mathbf{y}, \text{Cov}(\mathbf{x}, \mathbf{y})\text{Var}(\mathbf{y})^{-1}\mathbf{y}\} \\ &= \text{Cov}(\mathbf{y}, \mathbf{x}) - \text{Cov}(\mathbf{y}, \mathbf{y})\text{Var}(\mathbf{y})^{-1}\text{Cov}(\mathbf{x}, \mathbf{y})' \\ &= \text{Cov}(\mathbf{y}, \mathbf{x}) - \text{Var}(\mathbf{y})\text{Var}(\mathbf{y})^{-1}\text{Cov}(\mathbf{y}, \mathbf{x}) \\ &= 0 \end{aligned} \quad (2.26)$$

□

In the special case where  $\mathbf{x}, \mathbf{y}$  are jointly normally distributed, the fol-

lowing result holds:

**Lemma 2.5.3.** Let  $\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  where

$$\boldsymbol{\mu} = \begin{bmatrix} E(\mathbf{x}) \\ E(\mathbf{y}) \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \text{Var}(\mathbf{x}) & \text{Cov}(\mathbf{x}, \mathbf{y}) \\ \text{Cov}(\mathbf{y}, \mathbf{x}) & \text{Var}(\mathbf{y}) \end{bmatrix} \quad (2.27)$$

then the distribution of  $\mathbf{x}$  conditional on  $\mathbf{y}$  is also multivariate normal with mean

$$E(\mathbf{x}|\mathbf{y}) = E(\mathbf{x}) + \text{Cov}(\mathbf{x}, \mathbf{y})\text{Var}(\mathbf{y})^{-1}[\mathbf{y} - E(\mathbf{y})] \quad (2.28)$$

and variance matrix

$$\text{Var}(\mathbf{x}|\mathbf{y}) = \text{Var}(\mathbf{x}) - \text{Cov}(\mathbf{x}, \mathbf{y})\text{Var}(\mathbf{y})^{-1}\text{Cov}(\mathbf{y}, \mathbf{x}) \quad (2.29)$$

*Proof.* A proof can be found in Anderson & Moore (1979). □

Thus, in the Gaussian case, we have  $L(\mathbf{x}|\mathbf{y}) = E(\mathbf{x}|\mathbf{y})$  and  $\text{MSE}[L(\mathbf{x}|\mathbf{y})] = \text{Var}(\mathbf{x}|\mathbf{y})$ . This result is significant because the conditional expectation  $E(\mathbf{x}|\mathbf{y})$  is also the minimum mean-square estimator (MMSE) of  $\mathbf{x}$  given  $\mathbf{y}$  (Lehmann & Casella, 1990). Hence, in this case, the MMSLE is the MMSE. Furthermore, Corollary 2.5.2 now states that the conditional estimation error  $\mathbf{x} - E(\mathbf{x}|\mathbf{y})$  is uncorrelated with  $\mathbf{y}$  — these vectors are normally distributed and uncorrelated, hence they must be independent.

**Corollary 2.5.4.** Let  $\mathbf{x}, \mathbf{y}, \mathbf{z}$  be random vectors, and  $\mathbf{y}, \mathbf{z}$  uncorrelated. Then:

$$L(\mathbf{x}|\mathbf{y}, \mathbf{z}) = L(\mathbf{x}|\mathbf{y}) + \text{Cov}(\mathbf{x}, \mathbf{z})\text{Var}(\mathbf{z})^{-1}[\mathbf{z} - E(\mathbf{z})] \quad (2.30)$$

and its MSE as an estimator of  $\mathbf{x}$  is

$$\text{MSE}_{\mathbf{x}}[\text{L}(\mathbf{x}|\mathbf{y}, \mathbf{z})] = \text{MSE}_{\mathbf{x}}[\text{L}(\mathbf{x}|\mathbf{y})] - \text{Cov}(\mathbf{x}, \mathbf{z})\text{Var}(\mathbf{z})^{-1}\text{Cov}(\mathbf{z}, \mathbf{x}) \quad (2.31)$$

*Proof.* We apply Lemma 2.5.1 to the vectors  $\mathbf{x}$  and  $\begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix}$ :

$$\begin{aligned} & \text{L}(\mathbf{x}|\mathbf{y}, \mathbf{z}) \\ &= \text{E}(\mathbf{x}) + \text{Cov}\left(\mathbf{x}, \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix}\right) \text{Var}\begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix}^{-1} \left( \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} - \text{E}\begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} \right) \\ &= \text{E}(\mathbf{x}) + [\text{Cov}(\mathbf{x}, \mathbf{y}), \text{Cov}(\mathbf{x}, \mathbf{z})] \begin{bmatrix} \text{Var}(\mathbf{y})^{-1} & 0 \\ 0 & \text{Var}(\mathbf{z})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{y} - \text{E}(\mathbf{y}) \\ \mathbf{z} - \text{E}(\mathbf{z}) \end{bmatrix} \\ &= \text{E}(\mathbf{x}) + \text{Cov}(\mathbf{x}, \mathbf{y})\text{Var}(\mathbf{y})^{-1}[\mathbf{y} - \text{E}(\mathbf{y})] + \text{Cov}(\mathbf{x}, \mathbf{z})\text{Var}(\mathbf{z})^{-1}[\mathbf{z} - \text{E}(\mathbf{z})] \\ &= \text{L}(\mathbf{x}|\mathbf{y}) + \text{Cov}(\mathbf{x}, \mathbf{z})\text{Var}(\mathbf{z})^{-1}[\mathbf{z} - \text{E}(\mathbf{z})] \end{aligned} \quad (2.32)$$

Similarly for the MSE:

$$\begin{aligned} & \text{MSE}_{\mathbf{x}}[\text{L}(\mathbf{x}|\mathbf{y}, \mathbf{z})] \\ &= \text{Var}(\mathbf{x}) - \text{Cov}\left(\mathbf{x}, \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix}\right) \text{Var}\begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix}^{-1} \text{Cov}\left(\begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix}, \mathbf{x}\right) \\ &= \text{Var}(\mathbf{x}) - [\text{Cov}(\mathbf{x}, \mathbf{y}), \text{Cov}(\mathbf{x}, \mathbf{z})] \begin{bmatrix} \text{Var}(\mathbf{y})^{-1} & 0 \\ 0 & \text{Var}(\mathbf{z})^{-1} \end{bmatrix} \begin{bmatrix} \text{Cov}(\mathbf{y}, \mathbf{x}) \\ \text{Cov}(\mathbf{z}, \mathbf{x}) \end{bmatrix} \\ &= \text{Var}(\mathbf{x}) - \text{Cov}(\mathbf{x}, \mathbf{y})\text{Var}(\mathbf{y})^{-1}\text{Cov}(\mathbf{y}, \mathbf{x}) - \text{Cov}(\mathbf{x}, \mathbf{z})\text{Var}(\mathbf{z})^{-1}\text{Cov}(\mathbf{z}, \mathbf{x}) \\ &= \text{MSE}_{\mathbf{x}}[\text{L}(\mathbf{x}|\mathbf{y})] - \text{Cov}(\mathbf{x}, \mathbf{z})\text{Var}(\mathbf{z})^{-1}\text{Cov}(\mathbf{z}, \mathbf{x}) \end{aligned} \quad (2.33)$$

□

It follows that, in the Gaussian case, these results simplify to

$$\begin{aligned} E(\mathbf{x}|\mathbf{y}, \mathbf{z}) &= E(\mathbf{x}|\mathbf{y}) + \text{Cov}(\mathbf{x}, \mathbf{z})\text{Var}(\mathbf{z})^{-1}[\mathbf{z} - E(\mathbf{z})] \\ \text{Var}(\mathbf{x}|\mathbf{y}, \mathbf{z}) &= \text{Var}(\mathbf{x}|\mathbf{y}) - \text{Cov}(\mathbf{x}, \mathbf{z})\text{Var}(\mathbf{z})^{-1}\text{Cov}(\mathbf{z}, \mathbf{x}) \end{aligned} \quad (2.34)$$

### 2.5.3 Derivation

We will now derive the Kalman filter recursions. Let  $\mathbf{Y}_t = \{\mathbf{y}_1, \dots, \mathbf{y}_t\}$ ,  $t = 1, \dots, n$ . The KF obtains the one-step ahead MMSLE of the state vector  $\mathbf{a}_{t+1} = L(\boldsymbol{\alpha}_{t+1}|\mathbf{Y}_t)$  and its MSE  $\mathbf{P}_{t+1} = \text{MSE}_{\boldsymbol{\alpha}_{t+1}}(\mathbf{a}_{t+1})$ , for  $t = 1, \dots, n$ . In the Gaussian case, the errors  $\{\boldsymbol{\varepsilon}_t\}$  are normally distributed, all observations are normally distributed, hence conditional distributions of any subset are also normal. Lemma 2.5.3 implies  $\boldsymbol{\alpha}_{t+1}|\mathbf{Y}_t \sim N(\mathbf{a}_{t+1}, \mathbf{P}_{t+1})$ , thus  $\mathbf{a}_{t+1}$  is the one-step ahead MMSE of  $\boldsymbol{\alpha}_{t+1}$  and the quantities  $\mathbf{a}_{t+1}, \mathbf{P}_{t+1}$  are all we require to fully define the conditional distribution of  $\boldsymbol{\alpha}_{t+1}$  given past observations  $\mathbf{Y}_t$ .

Define the *innovations*:

$$\mathbf{v}_t \equiv \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t \quad (2.35)$$

Their expectation and variance are

$$\begin{aligned}
E(\mathbf{v}_t) &= E(\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \\
&= E[\mathbf{Z}_t(\boldsymbol{\alpha}_t - \mathbf{a}_t) + \mathbf{G}_t \boldsymbol{\varepsilon}_t] \\
&= \mathbf{Z}_t[E(\boldsymbol{\alpha}_t) - E(\mathbf{a}_t)] + \mathbf{G}_t E(\boldsymbol{\varepsilon}_t) \\
&= \mathbf{Z}_t[E(\boldsymbol{\alpha}_t) - E\{L(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1})\}] + \mathbf{G}_t E(\boldsymbol{\varepsilon}_t) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\text{Var}(\mathbf{v}_t) &= \text{Var}[\mathbf{Z}_t(\boldsymbol{\alpha}_t - \mathbf{a}_t) + \mathbf{G}_t \boldsymbol{\varepsilon}_t] \\
&= \mathbf{Z}_t \text{Var}(\boldsymbol{\alpha}_t - \mathbf{a}_t) \mathbf{Z}_t' + \mathbf{G}_t \text{Var}(\boldsymbol{\varepsilon}_t) \mathbf{G}_t' \\
&= \mathbf{Z}_t \text{MSE}(\mathbf{a}_t) \mathbf{Z}_t' + \mathbf{G}_t \mathbf{G}_t' \\
&\quad \quad \quad \boldsymbol{\alpha}_t \\
&= \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{G}_t \mathbf{G}_t' \\
&\equiv \mathbf{F}_t
\end{aligned} \tag{2.36}$$

Each innovation  $\mathbf{v}_t$  has the same dimension as the corresponding observation  $\mathbf{y}_t$  and is equal to the one-step-ahead prediction error for  $\mathbf{y}_t$ . In other words,  $\mathbf{v}_t$  is the “new” part of  $\mathbf{y}_t$ , which cannot be predicted using the past ( $\mathbf{Y}_{t-1}$ ):

$$\begin{aligned}
\text{Cov}(\mathbf{v}_t, \mathbf{Y}_{t-1}) &= E(\mathbf{v}_t \mathbf{Y}_{t-1}') - E(\mathbf{v}_t) E(\mathbf{Y}_{t-1})' \\
&= E[L(\mathbf{v}_t \mathbf{Y}_{t-1}' | \mathbf{Y}_{t-1})] \\
&= E[L\{\mathbf{Z}_t(\boldsymbol{\alpha}_t - \mathbf{a}_t) + \mathbf{G}_t \boldsymbol{\varepsilon}_t | \mathbf{Y}_{t-1}\} \mathbf{Y}_{t-1}'] \\
&= \mathbf{Z}_t E[\{L(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}) - L(\mathbf{a}_t | \mathbf{Y}_{t-1})\} \mathbf{Y}_{t-1}'] \\
&= \mathbf{Z}_t E[(\mathbf{a}_t - \mathbf{a}_t) \mathbf{Y}_{t-1}'] \\
&= 0
\end{aligned} \tag{2.37}$$

The span of  $\mathbf{Y}_t$  is the same as the span of  $\mathbf{Y}_{t-1}, \mathbf{v}_t$ , which are uncorrelated, so we can apply Corollary 2.5.4 to obtain



$$\begin{aligned}
\mathbf{a}_{t+1} &= \mathbf{L}(\boldsymbol{\alpha}_{t+1}|\mathbf{Y}_t) \\
&= \mathbf{L}(\boldsymbol{\alpha}_{t+1}|\mathbf{Y}_{t-1}, \mathbf{v}_t) \\
&= \mathbf{L}(\boldsymbol{\alpha}_{t+1}|\mathbf{Y}_{t-1}) + \text{Cov}(\boldsymbol{\alpha}_{t+1}, \mathbf{v}_t) \text{Var}(\mathbf{v}_t)^{-1} [\mathbf{v}_t - \mathbf{E}(\mathbf{v}_t)] \quad (2.38)
\end{aligned}$$

where the linear expectation on the RHS is

$$\begin{aligned}
\mathbf{L}(\boldsymbol{\alpha}_{t+1}|\mathbf{Y}_{t-1}) &= \mathbf{L}(\mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t | \mathbf{Y}_{t-1}) \\
&= \mathbf{T}_t \mathbf{L}(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}) + \mathbf{H}_t \mathbf{L}(\boldsymbol{\varepsilon}_t | \mathbf{Y}_{t-1}) \\
&= \mathbf{T}_t \mathbf{a}_t + \mathbf{H}_t \mathbf{E}(\boldsymbol{\varepsilon}_t) \\
&= \mathbf{T}_t \mathbf{a}_t \quad (2.39)
\end{aligned}$$

since  $\boldsymbol{\varepsilon}_t$  is independent of  $\mathbf{Y}_{t-1}$ . The final term is

$$\begin{aligned}
\text{Cov}(\boldsymbol{\alpha}_{t+1}, \mathbf{v}_t) \text{Var}(\mathbf{v}_t)^{-1} &= \text{Cov}[\mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t, \mathbf{Z}_t(\boldsymbol{\alpha}_t - \mathbf{a}_t) + \mathbf{G}_t \boldsymbol{\varepsilon}_t] \mathbf{F}_t^{-1} \\
&= [\mathbf{T}_t \text{Cov}(\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_t - \mathbf{a}_t) \mathbf{Z}_t' + \mathbf{H}_t \text{Var}(\boldsymbol{\varepsilon}_t) \mathbf{G}_t'] \mathbf{F}_t^{-1} \\
&= (\mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t \mathbf{G}_t') \mathbf{F}_t^{-1} \\
&\equiv \mathbf{K}_t \quad (2.40)
\end{aligned}$$

because

$$\begin{aligned}
\text{Cov}(\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_t - \mathbf{a}_t) &= \mathbf{E}[\boldsymbol{\alpha}_t(\boldsymbol{\alpha}_t - \mathbf{a}_t)'] \\
&= \mathbf{E}[(\boldsymbol{\alpha}_t - \mathbf{a}_t)(\boldsymbol{\alpha}_t - \mathbf{a}_t)'] + \mathbf{E}[\mathbf{a}_t(\boldsymbol{\alpha}_t - \mathbf{a}_t)'] \\
&= \text{MSE}_{\boldsymbol{\alpha}_t}(\mathbf{a}_t) + \mathbf{E}[\mathbf{L}\{\mathbf{a}_t(\boldsymbol{\alpha}_t - \mathbf{a}_t)' | \mathbf{Y}_{t-1}\}] \\
&= \mathbf{P}_t + \mathbf{E}[\mathbf{a}_t\{\mathbf{L}(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}) - \mathbf{L}(\mathbf{a}_t | \mathbf{Y}_{t-1})\}'] \\
&= \mathbf{P}_t + \mathbf{E}[\mathbf{a}_t\{\mathbf{a}_t - \mathbf{a}_t\}'] \\
&= \mathbf{P}_t \quad (2.41)
\end{aligned}$$

The quantity  $\mathbf{K}_t$  is known as the *Kalman gain*. Substituting back into (2.38) yields

$$\mathbf{a}_{t+1} = \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t \mathbf{v}_t \quad (2.42)$$

This is the updating equation for the filtered state estimate. Its MSE as a predictor of  $\boldsymbol{\alpha}_{t+1}$  is

$$\begin{aligned} \mathbf{P}_{t+1} &= \text{MSE}_{\boldsymbol{\alpha}_{t+1}}[\mathbf{L}(\boldsymbol{\alpha}_{t+1}|\mathbf{Y}_t)] \\ &= \text{MSE}_{\boldsymbol{\alpha}_{t+1}}[\mathbf{L}(\boldsymbol{\alpha}_{t+1}|\mathbf{Y}_{t-1}, \mathbf{v}_t)] \\ &= \text{MSE}_{\boldsymbol{\alpha}_{t+1}}[\mathbf{L}(\boldsymbol{\alpha}_{t+1}|\mathbf{Y}_{t-1})] - \text{Cov}(\boldsymbol{\alpha}_{t+1}, \mathbf{v}_t) \text{Var}(\mathbf{v}_t)^{-1} \text{Cov}(\mathbf{v}_t, \boldsymbol{\alpha}_{t+1}) \end{aligned} \quad (2.43)$$

The only term that has not already been evaluated is

$$\begin{aligned} \text{MSE}_{\boldsymbol{\alpha}_{t+1}}[\mathbf{L}(\boldsymbol{\alpha}_{t+1}|\mathbf{Y}_{t-1})] &= \text{MSE}_{\boldsymbol{\alpha}_{t+1}}(\mathbf{T}_t \mathbf{a}_t) \\ &= \text{Var}(\boldsymbol{\alpha}_{t+1} - \mathbf{T}_t \mathbf{a}_t) \\ &= \text{Var}[\mathbf{T}_t(\boldsymbol{\alpha}_t - \mathbf{a}_t) + \mathbf{H}_t \boldsymbol{\varepsilon}_t] \\ &= \mathbf{T}_t \text{Var}(\boldsymbol{\alpha}_t - \mathbf{a}_t) \mathbf{T}_t' + \mathbf{H}_t \text{Var}(\boldsymbol{\varepsilon}_t) \mathbf{H}_t' \\ &= \mathbf{T}_t \mathbf{P}_t \mathbf{T}_t' + \mathbf{H}_t \mathbf{H}_t' \end{aligned} \quad (2.44)$$

Substituting into (2.43) yields

$$\begin{aligned} \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{T}_t' + \mathbf{H}_t \mathbf{H}_t' - (\mathbf{K}_t \mathbf{F}_t) \mathbf{F}_t^{-1} (\mathbf{K}_t \mathbf{F}_t)' \\ &= \mathbf{T}_t \mathbf{P}_t \mathbf{T}_t' + \mathbf{H}_t \mathbf{H}_t' - \mathbf{K}_t \mathbf{F}_t \mathbf{K}_t' \end{aligned} \quad (2.45)$$

We define the matrix

$$\mathbf{L}_t \equiv \mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t \quad (2.46)$$

which allows us to write

$$\begin{aligned} \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{T}_t' + \mathbf{H}_t \mathbf{H}_t' - \mathbf{K}_t \mathbf{F}_t \mathbf{K}_t' \\ &= \mathbf{T}_t \mathbf{P}_t (\mathbf{L}_t' + \mathbf{Z}_t' \mathbf{K}_t) + \mathbf{H}_t \mathbf{H}_t' - (\mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t \mathbf{G}_t') \mathbf{F}_t^{-1} \mathbf{F}_t \mathbf{K}_t' \\ &= \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{H}_t (\mathbf{H}_t - \mathbf{K}_t \mathbf{G}_t)' \end{aligned} \quad (2.47)$$

Grouped together, equations (2.35), (2.36), (2.40), (2.46), (2.42) and (2.45) define the Kalman filter (KF) recursions:

$$\begin{aligned} \mathbf{v}_t &= \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t \\ \mathbf{F}_t &= \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{G}_t \mathbf{G}_t' \\ \mathbf{K}_t &= (\mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t \mathbf{G}_t') \mathbf{F}_t^{-1} \\ \mathbf{L}_t &= \mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t \\ \mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t \mathbf{v}_t \\ \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{H}_t (\mathbf{H}_t - \mathbf{K}_t \mathbf{G}_t)' \end{aligned} \quad (2.48)$$

The KF recursions are initialised with the unconditional mean and variance of  $\alpha_1$  ( $\mathbf{a}_1$  and  $\mathbf{P}_1$ , respectively) and run forwards for  $t = 1, \dots, n$ . The variance matrix  $\mathbf{P}_1$  often contains diffuse elements (de Jong, 1991; Durbin & Koopman, 2001). A numerical approach to modelling a time series with diffuse initial conditions is to define

$$\mathbf{P}_1 = \mathbf{P}_* + \kappa \mathbf{P}_\infty \quad (2.49)$$

where  $\mathbf{P}_*$ ,  $\mathbf{P}_\infty$  are symmetric matrices of the same dimensions as  $\mathbf{P}_1$ , and  $\kappa$  is a very large number (for example,  $\kappa = 10^6$ ).  $\mathbf{P}_*$  accounts for the variance of the non-diffuse components, whereas  $\mathbf{P}_\infty$  is a diagonal matrix

with unity values at the entries corresponding to the diffuse elements of  $\alpha_1$ , and zero values elsewhere.

This simple method produces an approximation to the KF recursions with diffuse initial conditions. Introducing large numbers into the recursions can potentially lead to numerical instability; Ansley & Kohn (1985) and de Jong (1991b) propose alternative approaches which overcome this problem.

#### 2.5.4 Cholesky Decomposition

The span of  $Y_{n-1}$  is the same as the span of  $v_1, \dots, v_{n-1}$ , thus  $v_n$  is uncorrelated with past innovations. Applying this argument recursively, we can deduce that all the innovations are pairwise uncorrelated. Furthermore, by repeated application of the KF recursions, we have

$$\begin{aligned} v_t &= y_t - Z_t a_t \\ &= y_t - Z_t L(\alpha_t | Y_{t-1}) \\ &= y_t - \sum_{j=1}^{t-1} W_j y_{t-j} \end{aligned} \tag{2.50}$$

for suitable weights  $W_j$ . Thus, each innovation  $v_t$  can be expressed as a linear combination of the observations up to time  $t$ . This implies that  $v = Ly$ , where  $L$  is a block-lower-triangular matrix and  $y, v$  are the stack vectors of observations and innovations, respectively. The matrix  $L$  is of full rank because it is block-triangular and has unit matrices along the main diagonal ( $W_o = I$ ), thus the linear transformation of the observations  $y$  into innovations  $v$  is non-singular. If we define the variance matrices  $\Sigma \equiv \text{Var}(y)$  and  $F \equiv \text{Var}(v)$ , we can write

$$\begin{aligned} F &= \text{Var}(Ly) = L\Sigma L' \\ \Leftrightarrow \Sigma &= L^{-1}F(L')^{-1} \end{aligned} \tag{2.51}$$

$$\Leftrightarrow \Sigma = QFQ' \tag{2.52}$$

where  $\mathbf{Q} = \mathbf{L}^{-1}$  is also a block-lower-triangular matrix. The matrix  $\mathbf{F} = \text{diag}\{\mathbf{F}_1, \dots, \mathbf{F}_t\}$  is block-diagonal, thus we can say that the KF implicitly computes the Cholesky decomposition of  $\mathbf{\Sigma}$ . This is the root of the computational efficiency of the KF. The linear estimation formulae require us to compute the variance matrix of the vectors we are conditioning on. Rather than computing MMSLEs by conditioning on the observations, we condition on the innovations, which have the same span (thus yield the same results), but are uncorrelated. This way, we avoid the computationally costly step of inverting the variance matrix of observations, and instead invert the block-diagonal matrix  $\mathbf{F}$ .

### 2.5.5 Model Estimation

Time series models in state space form can be estimated by maximizing the likelihood. Consider the likelihood function for the entire sample; by successive conditioning, we can write

$$\begin{aligned}\ell(\mathbf{y}_1, \dots, \mathbf{y}_n) &= \log f(\mathbf{y}_n, \mathbf{y}_{n-1}, \dots, \mathbf{y}_1) \\ &= \log \{f(\mathbf{y}_n | \mathbf{y}_{n-1}, \dots, \mathbf{y}_1) f(\mathbf{y}_{n-1} | \mathbf{y}_{n-2}, \dots, \mathbf{y}_1) \dots f(\mathbf{y}_1)\} \\ &= \sum_{t=1}^n \log f(\mathbf{y}_t | \mathbf{Y}_{t-1})\end{aligned}\tag{2.53}$$

where the final term is  $f(\mathbf{y}_1 | \mathbf{Y}_0) = f(\mathbf{y}_1)$ . The advantage of this formulation is that, in the Gaussian case, the conditional distribution of  $\mathbf{y}_t$  given  $\mathbf{Y}_{t-1}$  is also Gaussian, with expectation  $E(\mathbf{y}_t | \mathbf{Y}_{t-1}) = \mathbf{Z}_t \mathbf{a}_t$  and variance  $\text{Var}(\mathbf{y}_t | \mathbf{Y}_{t-1}) = \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{G}_t \mathbf{G}_t' = \mathbf{F}_t$ . Thus:

$$\begin{aligned}f(\mathbf{y}_t | \mathbf{Y}_{t-1}) &= (2\pi)^{-p/2} |\mathbf{F}_t|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t)' \mathbf{F}_t^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \right\} \\ &= (2\pi)^{-p/2} |\mathbf{F}_t|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \right\}\end{aligned}\tag{2.54}$$

Substituting into (2.53) yields

$$\ell(\mathbf{Y}_n) = -\frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n (\log |\mathbf{F}_t| + \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t) \quad (2.55)$$

This expression for the log-likelihood is known as the prediction error decomposition (Harvey, 1993). Given a set of parameter values, it enables us to compute the log-likelihood in a single forward pass of the Kalman filter.

## 2.6 The Kalman Filter Smoother

### 2.6.1 Fixed-Interval Smoothing

We now consider the problem known as *fixed-interval smoothing*: given the observations  $\mathbf{y}_1, \dots, \mathbf{y}_n$ , we want to find the MMSLE of the state  $\alpha_t$  and its MSE, conditional on the whole sample. We have:

$$\begin{aligned} \hat{\alpha}_t &= L(\alpha_t | \mathbf{Y}_n) \\ &= L(\alpha_t | \mathbf{Y}_{t-1}, \mathbf{v}_t, \dots, \mathbf{v}_n) \\ &= L(\alpha_t | \mathbf{Y}_{t-1}) + \text{Cov}(\alpha_t, \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix}) \text{Var} \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix}^{-1} \left( \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix} - E \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix} \right) \\ &= \mathbf{a}_t + [\text{Cov}(\alpha_t, \mathbf{v}_t) \dots \text{Cov}(\alpha_t, \mathbf{v}_n)] \begin{bmatrix} \mathbf{F}_t & & \\ & \ddots & \\ & & \mathbf{F}_n \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix} \\ &= \mathbf{a}_t + \sum_{s=t}^n \text{Cov}(\alpha_t, \mathbf{v}_s) \mathbf{F}_s^{-1} \mathbf{v}_s \end{aligned} \quad (2.56)$$

We will now prove the following lemma:

**Lemma 2.6.1.** *If  $s = t, t+1, \dots, n$  then*

$$\text{Cov}(\alpha_t, \mathbf{v}_s) = \mathbf{P}_t \underline{\mathbf{L}}'_{t,s-1} \mathbf{Z}'_s \quad (2.57)$$

$$\text{Cov}(\alpha_t, \alpha_s - \mathbf{a}_s) = \mathbf{P}_t \underline{\mathbf{L}}'_{t,s-1} \quad (2.58)$$

where we define

$$\underline{\mathbf{L}}'_{t,s-1} = \begin{cases} \mathbf{L}'_t \dots \mathbf{L}'_{s-1} & s = t+2, t+3, \dots \\ \mathbf{L}'_t & s = t+1 \\ \mathbf{I} & s = t \end{cases} \quad (2.59)$$

*Proof.* By induction on  $s$ . For  $s = t$ , we have already proved the result  $\text{Cov}(\alpha_t, \alpha_t - \mathbf{a}_t) = \mathbf{P}_t$ . We also have

$$\begin{aligned} \text{Cov}(\alpha_t, \mathbf{v}_t) &= \text{Cov}[\alpha_t, \mathbf{Z}_t(\alpha_t - \mathbf{a}_t) + \mathbf{G}_t \epsilon_t] \\ &= \text{Cov}(\alpha_t, \alpha_t - \mathbf{a}_t) \mathbf{Z}'_t + \text{Cov}(\alpha_t, \epsilon_t) \mathbf{G}'_t \\ &= \mathbf{P}_t \mathbf{Z}'_t \end{aligned} \quad (2.60)$$

thus the statements hold. Assuming both statements are true for all integers up to  $s-1$ , the inductive step is

$$\begin{aligned} \text{Cov}[\alpha_t, \alpha_s - \mathbf{a}_s] &= \text{Cov}[\alpha_t, \mathbf{T}_{s-1}(\alpha_{s-1} - \mathbf{a}_{s-1}) + \mathbf{H}_{s-1} \epsilon_{s-1} - \mathbf{K}_{s-1} \mathbf{v}_{s-1}] \\ &= \text{Cov}[\alpha_t, \alpha_{s-1} - \mathbf{a}_{s-1}] \mathbf{T}'_{s-1} - \text{Cov}(\alpha_t, \mathbf{v}_{s-1}) \mathbf{K}'_{s-1} \\ &= \mathbf{P}_t \underline{\mathbf{L}}'_{t,s-2} \mathbf{T}'_{s-1} - \mathbf{P}_t \underline{\mathbf{L}}'_{t,s-2} \mathbf{Z}'_{s-1} \mathbf{K}'_{s-1} \\ &= \mathbf{P}_t \underline{\mathbf{L}}'_{t,s-2} (\mathbf{T}'_{s-1} - \mathbf{Z}'_{s-1} \mathbf{K}'_{s-1}) \\ &= \mathbf{P}_t \underline{\mathbf{L}}'_{t,s-2} \mathbf{L}'_{s-1} \\ &= \mathbf{P}_t \underline{\mathbf{L}}'_{t,s-1} \end{aligned} \quad (2.61)$$

which implies

$$\begin{aligned}
\text{Cov}(\boldsymbol{\alpha}_t, \mathbf{v}_s) &= \text{Cov}[\boldsymbol{\alpha}_t, \mathbf{Z}_s(\boldsymbol{\alpha}_s - \mathbf{a}_s) + \mathbf{G}_s \boldsymbol{\varepsilon}_s] \\
&= \text{Cov}[\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_s - \mathbf{a}_s] \mathbf{Z}_s' \\
&= \mathbf{P}_t \underline{\mathbf{L}}'_{t,s-1} \mathbf{Z}_s'
\end{aligned} \tag{2.62}$$

□

Substituting the expression from Lemma 2.6.1 into (2.56) yields

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_t &= \mathbf{a}_t + \mathbf{P}_t \sum_{s=t}^n \underline{\mathbf{L}}'_{t,s-1} \mathbf{Z}_s' \mathbf{F}_s^{-1} \mathbf{v}_s \\
&= \mathbf{a}_t + \mathbf{P}_t \mathbf{r}_{t-1}
\end{aligned} \tag{2.63}$$

where we define the quantity  $\mathbf{r}_t \equiv \sum_{s=t+1}^n \underline{\mathbf{L}}'_{t+1,s-1} \mathbf{Z}_s' \mathbf{F}_s^{-1} \mathbf{v}_s$  for  $t = 0, \dots, n-1$ . It is possible to calculate  $\mathbf{r}_t$  recursively, using the identity

$$\begin{aligned}
\mathbf{r}_{t-1} &= \sum_{s=t}^n \underline{\mathbf{L}}'_{t,s-1} \mathbf{Z}_s' \mathbf{F}_s^{-1} \mathbf{v}_s \\
&= \sum_{s=t+1}^n \underline{\mathbf{L}}'_{t,s-1} \mathbf{Z}_s' \mathbf{F}_s^{-1} \mathbf{v}_s + \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \\
&= \mathbf{L}_t' \sum_{s=t+1}^n \underline{\mathbf{L}}'_{t+1,s-1} \mathbf{Z}_s' \mathbf{F}_s^{-1} \mathbf{v}_s + \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \\
&= \mathbf{L}_t' \mathbf{r}_t + \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{v}_t
\end{aligned} \tag{2.64}$$

which is initialised with  $\mathbf{r}_n = 0$  and runs backwards for  $t = n, n-1, \dots, 1$ .

The MSE of  $\hat{\boldsymbol{\alpha}}_t$  as a predictor of  $\boldsymbol{\alpha}_t$  is



$$\begin{aligned}
& \text{MSE}_{\alpha_t}(\hat{\alpha}_t) \\
&= \text{MSE}_{\alpha_t}[\text{L}(\alpha_t | \mathbf{Y}_n)] \\
&= \text{MSE}_{\alpha_t}[\text{L}(\alpha_t | \mathbf{Y}_{t-1}, \mathbf{v}_t, \dots, \mathbf{v}_n)] \\
&= \text{MSE}_{\alpha_t}[\text{L}(\alpha_t | \mathbf{Y}_{t-1})] - \text{Cov}(\alpha_t, \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix}) \text{Var} \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix}^{-1} \text{Cov}(\begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix}, \alpha_t) \\
&= \mathbf{P}_t - [\text{Cov}(\alpha_t, \mathbf{v}_t) \dots \text{Cov}(\alpha_t, \mathbf{v}_n)] \begin{bmatrix} \mathbf{F}_t & & \\ & \ddots & \\ & & \mathbf{F}_n \end{bmatrix}^{-1} \begin{bmatrix} \text{Cov}(\mathbf{v}_t, \alpha_t) \\ \vdots \\ \text{Cov}(\mathbf{v}_n, \alpha_t) \end{bmatrix} \\
&= \mathbf{P}_t - \sum_{s=t}^n \text{Cov}(\alpha_t, \mathbf{v}_s) \mathbf{F}_s^{-1} \text{Cov}(\mathbf{v}_s, \alpha_t) \\
&= \mathbf{P}_t - \sum_{s=t}^n \mathbf{P}_t \underline{\mathbf{L}}'_{t,s-1} \mathbf{Z}'_s \mathbf{F}_s^{-1} \mathbf{Z}_s \underline{\mathbf{L}}_{t,s-1} \mathbf{P}_t \\
&= \mathbf{P}_t - \mathbf{P}_t \mathbf{N}_{t-1} \mathbf{P}_t' \tag{2.65}
\end{aligned}$$

where we define the quantity  $\mathbf{N}_t \equiv \sum_{s=t+1}^n \underline{\mathbf{L}}'_{t+1,s-1} \mathbf{Z}'_s \mathbf{F}_s^{-1} \mathbf{Z}_s \underline{\mathbf{L}}_{t+1,s-1}$  for  $t = 0, \dots, n-1$ . The matrix  $\mathbf{N}_t$  is calculated recursively, using the identity

$$\begin{aligned}
\mathbf{N}_{t-1} &= \sum_{s=t}^n \underline{\mathbf{L}}'_{t,s-1} \mathbf{Z}'_s \mathbf{F}_s^{-1} \mathbf{Z}_s \underline{\mathbf{L}}_{t,s-1} \\
&= \sum_{s=t+1}^n \underline{\mathbf{L}}'_{t,s-1} \mathbf{Z}'_s \mathbf{F}_s^{-1} \mathbf{Z}_s \underline{\mathbf{L}}_{t,s-1} + \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \\
&= \mathbf{L}'_t \left( \sum_{s=t+1}^n \underline{\mathbf{L}}'_{t+1,s-1} \mathbf{Z}'_s \mathbf{F}_s^{-1} \mathbf{Z}_s \underline{\mathbf{L}}_{t+1,s-1} \right) \mathbf{L}_t + \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \\
&= \mathbf{L}'_t \mathbf{N}_t \mathbf{L}_t + \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \tag{2.66}
\end{aligned}$$

which is initialised with  $\mathbf{N}_n = 0$  and runs backwards for  $t = n, n-1, \dots, 1$ . The quantity  $\mathbf{N}_t$  is also the variance of  $\mathbf{r}_t$ :

$$\begin{aligned}
\text{Var}(\mathbf{r}_t) &= \text{Var} \left( \sum_{s=t+1}^n \underline{\mathbf{L}}'_{t+1,s-1} \mathbf{Z}'_s \mathbf{F}_s^{-1} \mathbf{v}_s \right) \\
&= \sum_{s=t+1}^n \underline{\mathbf{L}}'_{t+1,s-1} \mathbf{Z}'_s \mathbf{F}_s^{-1} \text{Var}(\mathbf{v}_s) \mathbf{F}_s^{-1} \mathbf{Z}_s \underline{\mathbf{L}}_{t+1,s-1} \\
&= \sum_{s=t+1}^n \underline{\mathbf{L}}'_{t+1,s-1} \mathbf{Z}'_s \mathbf{F}_s^{-1} \mathbf{Z}_s \underline{\mathbf{L}}_{t+1,s-1} \\
&= \mathbf{N}_t
\end{aligned} \tag{2.67}$$

Grouped together, equations (2.64), (2.66), (2.63) and (2.65) define the basic form of the Kalman filter smoother (KFS) recursions (de Jong, 1988; de Jong 1989; Kohn & Ansley, 1989):

$$\begin{aligned}
\mathbf{r}_{t-1} &= \mathbf{L}'_t \mathbf{r}_t + \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{v}_t \\
\mathbf{N}_{t-1} &= \mathbf{L}'_t \mathbf{N}_t \mathbf{L}_t + \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \\
\hat{\boldsymbol{\alpha}}_t &= \mathbf{a}_t + \mathbf{P}_t \mathbf{r}_{t-1} \\
\text{MSE}_{\hat{\boldsymbol{\alpha}}_t}(\hat{\boldsymbol{\alpha}}_t) &= \mathbf{P}_t - \mathbf{P}_t \mathbf{N}_{t-1} \mathbf{P}'_t
\end{aligned} \tag{2.68}$$

The KFS recursions use the output from the forward pass of the KF. They are initialised with  $\mathbf{r}_n = 0, \mathbf{N}_n = 0$  and run backwards for  $t = n, n-1, \dots, 1$ . The filtered state estimates and their MSE, the innovations and their variance, and the Kalman gain ( $\mathbf{a}_t, \mathbf{P}_t, \mathbf{v}_t, \mathbf{F}_t, \mathbf{K}_t$ , respectively), need to be stored during the forward pass. The terms  $\mathbf{L}_t$  can be recovered from  $\mathbf{K}_t$ , so we prefer to store the  $q \times p$  matrices  $\mathbf{K}_t$  rather than the  $q \times q$  matrices  $\mathbf{L}_t$ . This is because, in practice, the length of the observation vector ( $p$ ) tends to be smaller than the length of the state vector ( $q$ ), sometimes considerably so.

### 2.6.2 Disturbance Smoothing

We will now derive smoothed estimates of the error terms conditional on the whole sample. This is known as disturbance smoothing (Koopman, 1993). The MMSLE of  $\boldsymbol{\varepsilon}_t$  is

$$\begin{aligned}
\hat{\epsilon}_t &= \mathbf{L}(\epsilon_t | \mathbf{Y}_n) \\
&= \mathbf{L}(\epsilon_t | \mathbf{Y}_{t-1}, \mathbf{v}_t, \dots, \mathbf{v}_n) \\
&= \mathbf{L}(\epsilon_t | \mathbf{Y}_{t-1}) + \text{Cov}(\epsilon_t, \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix}) \text{Var} \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix}^{-1} \left( \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix} - \mathbf{E} \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix} \right) \\
&= [\text{Cov}(\epsilon_t, \mathbf{v}_t) \dots \text{Cov}(\epsilon_t, \mathbf{v}_n)] \begin{bmatrix} \mathbf{F}_t & & \\ & \ddots & \\ & & \mathbf{F}_n \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix} \\
&= \sum_{s=t}^n \text{Cov}(\epsilon_t, \mathbf{v}_s) \mathbf{F}_s^{-1} \mathbf{v}_s \tag{2.69}
\end{aligned}$$

The first covariance term, for  $s = t$ , is

$$\begin{aligned}
\text{Cov}(\epsilon_t, \mathbf{v}_t) &= \text{Cov}(\epsilon_t, \mathbf{Z}_t(\alpha_t - \mathbf{a}_t) + \mathbf{G}_t \epsilon_t) \\
&= \text{Cov}(\epsilon_t, \epsilon_t) \mathbf{G}_t' \\
&= \mathbf{G}_t' \tag{2.70}
\end{aligned}$$

For  $s = t + 1, \dots, n$ , we have

$$\begin{aligned}
\text{Cov}(\epsilon_t, \mathbf{v}_s) &= \text{Cov}[\alpha_t, \mathbf{Z}_s(\alpha_s - \mathbf{a}_s) + \mathbf{G}_s \epsilon_s] \\
&= \text{Cov}[\epsilon_t, \alpha_s - \mathbf{a}_s] \mathbf{Z}_s' \tag{2.71}
\end{aligned}$$

The remaining covariance term can now be evaluated:

$$\begin{aligned}
& \text{Cov}[\boldsymbol{\varepsilon}_t, \boldsymbol{\alpha}_s - \mathbf{a}_s] \\
&= \text{Cov}[\boldsymbol{\varepsilon}_t, \mathbf{T}_{s-1}(\boldsymbol{\alpha}_{s-1} - \mathbf{a}_{s-1}) + \mathbf{H}_{s-1}\boldsymbol{\varepsilon}_{s-1} - \mathbf{K}_{s-1}\mathbf{v}_{s-1}] \\
&= \text{Cov}[\boldsymbol{\varepsilon}_t, \boldsymbol{\alpha}_{s-1} - \mathbf{a}_{s-1}]\mathbf{T}_{s-1}' - \text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{v}_{s-1})\mathbf{K}_{s-1}' \\
&= \text{Cov}[\boldsymbol{\varepsilon}_t, \boldsymbol{\alpha}_{s-1} - \mathbf{a}_{s-1}]\mathbf{T}_{s-1}' - \text{Cov}[\boldsymbol{\varepsilon}_t, \boldsymbol{\alpha}_{s-1} - \mathbf{a}_{s-1}]\mathbf{Z}_{s-1}'\mathbf{K}_{s-1}' \\
&= \text{Cov}[\boldsymbol{\varepsilon}_t, \boldsymbol{\alpha}_{s-1} - \mathbf{a}_{s-1}]\mathbf{L}_{s-1}' \\
&= \vdots \\
&= \text{Cov}[\boldsymbol{\varepsilon}_t, \boldsymbol{\alpha}_{t+1} - \mathbf{a}_{t+1}]\mathbf{L}_{t+1,s-1}' \\
&= \text{Cov}[\boldsymbol{\varepsilon}_t, \mathbf{T}_t(\boldsymbol{\alpha}_t - \mathbf{a}_t + \mathbf{H}_t\boldsymbol{\varepsilon}_t - \mathbf{K}_t\mathbf{v}_t)]\mathbf{L}_{t+1,s-1}' \\
&= [\text{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_t)\mathbf{H}_t' - \text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{v}_t)\mathbf{K}_t']\mathbf{L}_{t+1,s-1}' \\
&= (\mathbf{H}_t' - \mathbf{G}_t\mathbf{K}_t')\mathbf{L}_{t+1,s-1}' \tag{2.72}
\end{aligned}$$

Hence, for  $s = t + 1, \dots, n$ , we can write

$$\text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{v}_s) = (\mathbf{H}_t' - \mathbf{G}_t'\mathbf{K}_t')\mathbf{L}_{t+1,s-1}'\mathbf{Z}_s' \tag{2.73}$$

Substituting into (2.69) yields

$$\begin{aligned}
\hat{\boldsymbol{\varepsilon}}_t &= \text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{v}_t)\mathbf{F}_t^{-1}\mathbf{v}_t + \sum_{s=t+1}^n \text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{v}_s)\mathbf{F}_s^{-1}\mathbf{v}_s \\
&= \mathbf{G}_t'\mathbf{F}_t^{-1}\mathbf{v}_t + \sum_{s=t+1}^n (\mathbf{H}_t' - \mathbf{G}_t'\mathbf{K}_t')\mathbf{L}_{t+1,s-1}'\mathbf{Z}_s'\mathbf{F}_s^{-1}\mathbf{v}_s \\
&= \mathbf{G}_t'\mathbf{F}_t^{-1}\mathbf{v}_t + (\mathbf{H}_t' - \mathbf{G}_t'\mathbf{K}_t')\mathbf{r}_t \\
&= \mathbf{G}_t'(\mathbf{F}_t^{-1}\mathbf{v}_t - \mathbf{K}_t'\mathbf{r}_t) + \mathbf{H}_t'\mathbf{r}_t \\
&= \mathbf{G}_t'\mathbf{u}_t + \mathbf{H}_t'\mathbf{r}_t \tag{2.74}
\end{aligned}$$

where we define

$$\mathbf{u}_t \equiv \mathbf{F}_t^{-1}\mathbf{v}_t - \mathbf{K}_t'\mathbf{r}_t \tag{2.75}$$

for  $t = 1, \dots, n$ . The quantities  $\mathbf{u}_t$  are known as *smoothations* and have the same dimension as the observations  $\mathbf{y}_t$ . The variance of the smoothations is

$$\begin{aligned}
\text{Var}(\mathbf{u}_t) &= \text{Var}(\mathbf{F}_t^{-1}\mathbf{v}_t + \mathbf{K}_t'\mathbf{r}_t) \\
&= \mathbf{F}_t^{-1}\text{Var}(\mathbf{v}_t)\mathbf{F}_t^{-1} + \mathbf{K}_t'\text{Var}(\mathbf{r}_t)\mathbf{K}_t \\
&= \mathbf{F}_t^{-1} + \mathbf{K}_t'\mathbf{N}_t\mathbf{K}_t \\
&\equiv \mathbf{M}_t
\end{aligned} \tag{2.76}$$

It is possible to simplify the calculations by expressing  $\mathbf{r}_{t-1}$  as a function of  $\mathbf{u}_t$ :

$$\begin{aligned}
\mathbf{r}_{t-1} &= \mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{v}_t + \mathbf{L}_t'\mathbf{r}_t \\
&= \mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{v}_t + (\mathbf{T}_t' - \mathbf{Z}_t'\mathbf{K}_t')\mathbf{r}_t \\
&= \mathbf{Z}_t'(\mathbf{F}_t^{-1}\mathbf{v}_t - \mathbf{K}_t'\mathbf{r}_t) + \mathbf{T}_t'\mathbf{r}_t \\
&= \mathbf{Z}_t'\mathbf{u}_t + \mathbf{T}_t'\mathbf{r}_t
\end{aligned} \tag{2.77}$$

The MSE of  $\hat{\boldsymbol{\varepsilon}}_t$  is

$$\begin{aligned}
& \text{MSE}_{\boldsymbol{\varepsilon}_t}(\hat{\boldsymbol{\varepsilon}}_t) \\
&= \text{MSE}_{\boldsymbol{\varepsilon}_t}[\text{L}(\boldsymbol{\varepsilon}_t | \mathbf{Y}_n)] \\
&= \text{MSE}_{\boldsymbol{\varepsilon}_t}[\text{L}(\boldsymbol{\varepsilon}_t | \mathbf{Y}_{t-1}, \mathbf{v}_t, \dots, \mathbf{v}_n)] \\
&= \text{MSE}_{\boldsymbol{\varepsilon}_t}[\text{L}(\boldsymbol{\varepsilon}_t | \mathbf{Y}_{t-1})] - \text{Cov}(\boldsymbol{\varepsilon}_t, \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix}) \text{Var} \begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix}^{-1} \text{Cov}(\begin{bmatrix} \mathbf{v}_t \\ \vdots \\ \mathbf{v}_n \end{bmatrix}, \boldsymbol{\alpha}_t) \\
&= \text{Var}(\boldsymbol{\varepsilon}_t) - [\text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{v}_t) \dots \text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{v}_n)] \begin{bmatrix} \mathbf{F}_t & & \\ & \ddots & \\ & & \mathbf{F}_n \end{bmatrix}^{-1} \begin{bmatrix} \text{Cov}(\mathbf{v}_t, \boldsymbol{\varepsilon}_t) \\ \vdots \\ \text{Cov}(\mathbf{v}_n, \boldsymbol{\varepsilon}_t) \end{bmatrix} \\
&= \mathbf{I} - \sum_{s=t}^n \text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{v}_s) \mathbf{F}_s^{-1} \text{Cov}(\mathbf{v}_s, \boldsymbol{\varepsilon}_t) \\
&= \mathbf{I} - \text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{v}_t) \mathbf{F}_t^{-1} \text{Cov}(\mathbf{v}_t, \boldsymbol{\varepsilon}_t) - \sum_{s=t+1}^n \text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{v}_s) \mathbf{F}_s^{-1} \text{Cov}(\mathbf{v}_s, \boldsymbol{\varepsilon}_t) \\
&= \mathbf{I} - \mathbf{G}_t' \mathbf{F}_t^{-1} \mathbf{G}_t \\
&\quad - \sum_{s=t+1}^n (\mathbf{H}_t' - \mathbf{G}_t' \mathbf{K}_t') \underline{\mathbf{L}}_{t+1, s-1}' \mathbf{Z}_s' \mathbf{F}_s^{-1} \mathbf{Z}_s \underline{\mathbf{L}}_{t+1, s-1} (\mathbf{H}_t - \mathbf{K}_t \mathbf{G}_t) \\
&= \mathbf{I} - \mathbf{G}_t' \mathbf{F}_t^{-1} \mathbf{G}_t - (\mathbf{H}_t' - \mathbf{G}_t' \mathbf{K}_t') \mathbf{N}_t (\mathbf{H}_t - \mathbf{K}_t \mathbf{G}_t) \tag{2.78}
\end{aligned}$$

Grouped together, equations (2.75), (2.76), (2.77), along with (2.66), define the extended KFS recursions (de Jong & Penzer, 1998):

$$\begin{aligned}
\mathbf{u}_t &= \mathbf{F}_t^{-1} \mathbf{v}_t - \mathbf{K}_t' \mathbf{r}_t \\
\mathbf{M}_t &= \mathbf{F}_t^{-1} + \mathbf{K}_t' \mathbf{N}_t \mathbf{K}_t \\
\mathbf{r}_{t-1} &= \mathbf{Z}_t' \mathbf{u}_t + \mathbf{T}_t' \mathbf{r}_t \\
\mathbf{N}_{t-1} &= \mathbf{L}_t' \mathbf{N}_t \mathbf{L}_t + \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t \tag{2.79}
\end{aligned}$$

which are initialised with  $\mathbf{r}_n = 0$ ,  $\mathbf{N}_n = 0$ , as before.

## Chapter 3

# Interpolated Estimates

### 3.1 Introduction

In the analysis of time series, it is often useful to treat an observation as missing and estimate it from the rest of the data. The resulting estimates, which we will refer to as *interpolated estimates*, have several applications, such as the detection of unusual observations (de Jong & Penzer, 1998; Proietti, 2003; Penzer, 2007). A simple, though computationally inefficient technique for obtaining interpolated estimates is to perform a pass of the filtering and smoothing recursions, omitting a single observation  $\mathbf{y}_t$  (Brockwell & Davis, 1996). The resulting smoothed estimates of the state and disturbance vectors ( $\hat{\boldsymbol{\alpha}}_t^*$  and  $\hat{\boldsymbol{\varepsilon}}_t^*$ , respectively) will be the linear expectations of these quantities conditional on the entire sample apart from  $\mathbf{y}_t$ . Thus, the interpolated estimate of  $\mathbf{y}_t$  can be evaluated as  $\mathbf{Z}_t\hat{\boldsymbol{\alpha}}_t^* + \mathbf{G}_t\hat{\boldsymbol{\varepsilon}}_t^*$ . This procedure, including the forward and backward pass of the recursions, needs to be repeated for each interpolated estimate.

A much more elegant method is to use the quantities  $\mathbf{M}_t, \mathbf{u}_t$  from the Kalman filter smoother recursions to compute the interpolated estimates directly. This is far more efficient, since we only require a single forward and backward pass of the KFS to obtain interpolated estimates for the whole sample. However, a general and theoretically sound proof of the fundamental result linking the KFS output with interpolated estimates does not appear in time series literature. Existing proofs either consider

only the special case of the SSF with uncorrelated measurement and transition errors (de Jong, 1989) or are insufficiently formal (de Jong & Penzer, 1998).

In this chapter, we establish a novel proof of the key result for the interpolated estimates of the observation vector (Theorem 3.2.1) and use a similar argument to construct the corresponding estimate of the state vector (Theorem 3.2.3). In addition, we consider the problem of constructing *partially interpolated estimates* of  $\mathbf{y}_t$  and  $\boldsymbol{\alpha}_t$ , that is, interpolated estimates conditional on the data available at time  $m$ , where  $t < m < n$ . We establish formulae for these in Theorems 3.3.2 and 3.3.3.

## 3.2 Interpolation

### 3.2.1 Introduction

Consider the time series  $\{\mathbf{y}_t\}$  in the general SSF (de Jong, 1991):

$$\begin{aligned}\mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{G}_t \boldsymbol{\varepsilon}_t \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t \quad , \quad t = 1, \dots, n\end{aligned}\tag{3.1}$$

where  $\{\boldsymbol{\varepsilon}_t\} \sim \text{WN}(\mathbf{0}, \sigma^2 \mathbf{I}_r)$  and  $\boldsymbol{\alpha}_1$  has mean  $\mathbf{a}_1$  and variance  $\mathbf{P}_1$ . We define the punctured space  $\mathbf{Y}_t^{(i)} \equiv \{\mathbf{y}_1, \dots, \mathbf{y}_{i-1}, \mathbf{y}_{i+1}, \dots, \mathbf{y}_t\}$ . We will now proceed to establish formulae for the interpolated MMSLE of the state and observation vectors, which we define as

$$\hat{\boldsymbol{\alpha}}_t \equiv \text{L}(\boldsymbol{\alpha}_t | \mathbf{Y}_n^{(t)})\tag{3.2}$$

$$\hat{\mathbf{y}}_t \equiv \text{L}(\mathbf{y}_t | \mathbf{Y}_n^{(t)})\tag{3.3}$$

### 3.2.2 Observation Estimate

We establish the following theorem for the general SSF:



**Theorem 3.2.1.** *The interpolated MMSLE of observation  $\mathbf{y}_t$  is given by*

$$\hat{\mathbf{y}}_t = \mathbf{y}_t - \mathbf{M}_t^{-1} \mathbf{u}_t \quad (3.4)$$

*and its mean square error is  $\mathbf{M}_t^{-1}$ , where  $\mathbf{u}_t, \mathbf{M}_t$  are obtained from the extended KFS recursions.*

A special case of Theorem 3.2.1 is proved by de Jong (1989); it applies to the SSF where the measurement and transition errors are uncorrelated ( $\mathbf{G}_t \mathbf{H}_t' = \mathbf{0}$ ). The general result appears in de Jong & Penzer (1998), but the outline proof provided is not as rigorous as that of the special case. Their argument relies on a particular representation of the KF and KFS recursions:

$$\left. \begin{array}{l} \mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t \\ \mathbf{a}_{t+1} = \mathbf{K}_t \mathbf{y}_t + \mathbf{L}_t \mathbf{a}_t \end{array} \right\} \Leftrightarrow \begin{bmatrix} \mathbf{v}_t \\ \mathbf{a}_{t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\mathbf{Z}_t \\ \mathbf{K}_t & \mathbf{L}_t \end{bmatrix} \begin{bmatrix} \mathbf{y}_t \\ \mathbf{a}_t \end{bmatrix} \quad (3.5)$$

$$\left. \begin{array}{l} \mathbf{u}_t = \mathbf{F}_t^{-1} \mathbf{v}_t - \mathbf{K}_t' \mathbf{r}_t \\ \mathbf{r}_{t-1} = \mathbf{Z}_t \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{L}_t' \mathbf{r}_t \end{array} \right\} \Leftrightarrow \begin{bmatrix} -\mathbf{u}_t \\ \mathbf{r}_{t-1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\mathbf{Z}_t \\ \mathbf{K}_t & \mathbf{L}_t \end{bmatrix}' \begin{bmatrix} -\mathbf{F}_t^{-1} \mathbf{v}_t \\ \mathbf{r}_t \end{bmatrix} \quad (3.6)$$

Thus, expressed in this form, the operations at each step of the smoother can be thought of as the transpose of the operations at the corresponding step of the filter. De Jong & Penzer extend this argument to the overall output of the filter and smoother in stack vector form. Using the notation from the technical introduction, if the effect of the KF recursions can be summarised as  $\mathbf{v} = \mathbf{L}\mathbf{y}$ , then the equivalent form for the KFS recursions is  $-\mathbf{u} = \mathbf{L}'(-\mathbf{F}^{-1}\mathbf{v})$ , where  $\mathbf{u}$  is the stack vector of  $\{\mathbf{u}_t\}$ . Thus:

$$\begin{aligned}
\mathbf{u} &= \mathbf{L}' \mathbf{F}^{-1} \mathbf{v} \\
&= \mathbf{L}' \mathbf{F}^{-1} \mathbf{L} \mathbf{y} \\
&= \boldsymbol{\Sigma} \mathbf{y}
\end{aligned} \tag{3.7}$$

This intermediate result, which is equivalent to expression (3.22) in our proof, is the most important step in establishing the main result.

We will now construct a more rigorous proof of Theorem (3.2.1).

*Proof.* Let  $\mathbf{x}_t$  be the *signal* at time  $t$ , the observation minus the associated measurement noise, defined as  $\mathbf{x}_t \equiv \mathbf{Z}_t \boldsymbol{\alpha}_t$ . The smoothed estimate of the signal is

$$\begin{aligned}
\hat{\mathbf{x}}_t &= \mathbf{L}(\mathbf{x}_t | \mathbf{Y}_n) \\
&= \mathbf{Z}_t \mathbf{L}(\boldsymbol{\alpha}_t | \mathbf{Y}_n) \\
&= \mathbf{Z}_t \hat{\boldsymbol{\alpha}}_t
\end{aligned} \tag{3.8}$$

The smoothed estimate of the measurement error  $\mathbf{G}_t \boldsymbol{\varepsilon}_t$  is the difference between the observation  $\mathbf{y}_t$  and the smoothed signal estimate  $\hat{\mathbf{x}}_t$ . It can be expressed in terms of the KFS output:

$$\begin{aligned}
\mathbf{y}_t - \hat{\mathbf{x}}_t &= \mathbf{Z}_t \mathbf{a}_t + \mathbf{v}_t - \mathbf{Z}_t \hat{\boldsymbol{\alpha}}_t && \text{(defn. of } \mathbf{v}_t) \\
&= \mathbf{Z}_t \mathbf{a}_t + \mathbf{v}_t - \mathbf{Z}_t (\mathbf{a}_t + \mathbf{P}_t \mathbf{r}_{t-1}) \\
&= \mathbf{v}_t - \mathbf{Z}_t \mathbf{P}_t \mathbf{r}_{t-1} \\
&= \mathbf{F}_t (\mathbf{u}_t + \mathbf{K}_t' \mathbf{r}_t) - \mathbf{Z}_t \mathbf{P}_t (\mathbf{Z}_t' \mathbf{u}_t + \mathbf{T}_t' \mathbf{r}_t) && \text{(defn. of } \mathbf{u}_t, \mathbf{r}_t) \\
&= (\mathbf{F}_t - \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t') \mathbf{u}_t + (\mathbf{F}_t \mathbf{K}_t' - \mathbf{Z}_t \mathbf{P}_t \mathbf{T}_t') \mathbf{r}_t
\end{aligned} \tag{3.9}$$

We can simplify the matrices in this expression by considering the definitions of  $\mathbf{F}_t$  and  $\mathbf{K}_t$ :

$$\begin{aligned}\mathbf{F}_t &= \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{G}_t \mathbf{G}_t' \\ \Rightarrow \mathbf{G}_t \mathbf{G}_t' &= \mathbf{F}_t - \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t'\end{aligned}\tag{3.10}$$

$$\begin{aligned}\mathbf{K}_t &= (\mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t \mathbf{G}_t') \mathbf{F}_t^{-1} \\ \Rightarrow \mathbf{K}_t' &= \mathbf{F}_t^{-1} (\mathbf{Z}_t \mathbf{P}_t \mathbf{T}_t' + \mathbf{G}_t \mathbf{H}_t') \\ \Rightarrow \mathbf{G}_t \mathbf{H}_t' &= \mathbf{F}_t \mathbf{K}_t' - \mathbf{Z}_t \mathbf{P}_t \mathbf{T}_t'\end{aligned}\tag{3.11}$$

These expressions allow us to write

$$\mathbf{y}_t - \hat{\mathbf{x}}_t = \mathbf{G}_t \mathbf{G}_t' \mathbf{u}_t + \mathbf{G}_t \mathbf{H}_t' \mathbf{r}_t\tag{3.12}$$

Note that the RHS is equal to  $\mathbf{G}_t \hat{\boldsymbol{\varepsilon}}_t$ , where  $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{L}(\boldsymbol{\varepsilon}_t | Y_n)$  are the smoothed estimates of the errors, as shown in (2.74). De Jong (1989) establishes an analogous result to (3.12) for the special case where contemporaneous measurement and transition errors are uncorrelated (that is,  $\mathbf{G}_t \mathbf{H}_t' = 0$ ).

Now suppose that  $\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\varepsilon}, \mathbf{x}$  and  $\hat{\mathbf{x}}$  are (respectively) the stacked vectors of  $\mathbf{y}_t, \boldsymbol{\alpha}_t, \boldsymbol{\varepsilon}_t, \mathbf{x}_t$  and  $\hat{\mathbf{x}}_t$  for  $t = 1, \dots, n$ . We also define the block-diagonal matrices  $\underline{\mathbf{G}} \equiv \text{diag}\{\mathbf{G}_1, \dots, \mathbf{G}_n\}$  and  $\underline{\mathbf{Z}} \equiv \text{diag}\{\mathbf{Z}_1, \dots, \mathbf{Z}_n\}$ . The measurement equation can be written in stack vector form as  $\mathbf{y} = \mathbf{x} + \underline{\mathbf{G}}\boldsymbol{\varepsilon} = \underline{\mathbf{Z}}\boldsymbol{\alpha} + \underline{\mathbf{G}}\boldsymbol{\varepsilon}$ . Using the linear expectation formula, we have

$$\hat{\mathbf{x}} = \mathbf{L}(\mathbf{x} | \mathbf{y}) = \mathbf{E}(\mathbf{x}) + \text{Cov}(\mathbf{x}, \mathbf{y}) \text{Var}(\mathbf{y})^{-1} \{\mathbf{y} - \mathbf{E}(\mathbf{y})\}\tag{3.13}$$

Clearly,  $\mathbf{E}(\boldsymbol{\varepsilon}) = 0$  so  $\mathbf{E}(\mathbf{x}) = \mathbf{E}(\mathbf{y})$ . We can write the covariance matrix of  $\mathbf{x}$  and  $\mathbf{y}$  as

$$\begin{aligned}
\text{Cov}(\mathbf{x}, \mathbf{y}) &= \text{Cov}(\mathbf{y} - \underline{\mathbf{G}}\boldsymbol{\varepsilon}, \mathbf{y}) \\
&= \text{Var}(\mathbf{y}) - \underline{\mathbf{G}}\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{y}) \\
&= \text{Var}(\mathbf{y}) - \underline{\mathbf{G}}\text{Cov}(\boldsymbol{\varepsilon}, \underline{\mathbf{Z}}\boldsymbol{\alpha} + \underline{\mathbf{G}}\boldsymbol{\varepsilon}) \\
&= \text{Var}(\mathbf{y}) - \underline{\mathbf{G}}\text{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\alpha})\underline{\mathbf{Z}}' - \underline{\mathbf{G}}\text{Var}(\boldsymbol{\varepsilon})\underline{\mathbf{G}}' \\
&= \boldsymbol{\Sigma} - \underline{\mathbf{G}}\underline{\mathbf{S}}\underline{\mathbf{Z}}' - \underline{\mathbf{G}}\underline{\mathbf{G}}' \tag{3.14}
\end{aligned}$$

where  $\boldsymbol{\Sigma} = \text{Var}(\mathbf{y})$ , as before, and  $\underline{\mathbf{S}} \equiv \text{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\alpha})$ . Substituting back into (3.13), we have

$$\begin{aligned}
\hat{\mathbf{x}} &= \text{E}(\mathbf{y}) + (\boldsymbol{\Sigma} - \underline{\mathbf{G}}\underline{\mathbf{S}}\underline{\mathbf{Z}}' - \underline{\mathbf{G}}\underline{\mathbf{G}}')\boldsymbol{\Sigma}^{-1}\{\mathbf{y} - \text{E}(\mathbf{y})\} \\
&= \text{E}(\mathbf{y}) + \{\mathbf{y} - \text{E}(\mathbf{y})\} - (\underline{\mathbf{G}}\underline{\mathbf{S}}\underline{\mathbf{Z}}' + \underline{\mathbf{G}}\underline{\mathbf{G}}')\boldsymbol{\Sigma}^{-1}\{\mathbf{y} - \text{E}(\mathbf{y})\} \\
&= \mathbf{y} - (\underline{\mathbf{G}}\underline{\mathbf{S}}\underline{\mathbf{Z}}' + \underline{\mathbf{G}}\underline{\mathbf{G}}')\boldsymbol{\Sigma}^{-1}\{\mathbf{y} - \text{E}(\mathbf{y})\} \tag{3.15}
\end{aligned}$$

We can rearrange (3.15) as

$$\mathbf{y} - \hat{\mathbf{x}} = (\underline{\mathbf{G}}\underline{\mathbf{S}}\underline{\mathbf{Z}}' + \underline{\mathbf{G}}\underline{\mathbf{G}}')\boldsymbol{\Sigma}^{-1}\{\mathbf{y} - \text{E}(\mathbf{y})\} \tag{3.16}$$

Note that the LHS of (3.12) is the  $t^{\text{th}}$  vector component of the LHS of (3.16), thus the same must hold for the RHS. We will now evaluate  $\mathbf{w}_t$ , the  $t^{\text{th}}$  vector component of  $\boldsymbol{\Sigma}^{-1}\{\mathbf{y} - \text{E}(\mathbf{y})\}$ . For convenience of notation, we rearrange the stack vector of observations and write it in the form  $\mathbf{y} = [\mathbf{y}'_{\setminus t} \ \mathbf{y}'_t]'$ , where  $\mathbf{y}_{\setminus t}$  is the stack vector with observation  $\mathbf{y}_t$  omitted. Using standard block-matrix inversion results, we have

$$\begin{aligned}
& \Sigma^{-1}\{\mathbf{y} - \mathbf{E}(\mathbf{y})\} \\
&= \begin{bmatrix} \Sigma_{\setminus t} & \Sigma_{\setminus t, t} \\ \Sigma'_{\setminus t, t} & \Sigma_t \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_{\setminus t} - \mathbf{E}(\mathbf{y}_{\setminus t}) \\ \mathbf{y}_t - \mathbf{E}(\mathbf{y}_t) \end{bmatrix} \\
&= \begin{bmatrix} * & * \\ -\Sigma_{t|\setminus t}^{-1} \Sigma_{t, \setminus t} \Sigma_{\setminus t}^{-1} & \Sigma_{t|\setminus t}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{\setminus t} - \mathbf{E}(\mathbf{y}_{\setminus t}) \\ \mathbf{y}_t - \mathbf{E}(\mathbf{y}_t) \end{bmatrix} \\
&= \begin{bmatrix} * \\ \mathbf{w}_t \end{bmatrix} \tag{3.17}
\end{aligned}$$

where  $\Sigma_t = \mathbf{Var}(\mathbf{y}_t)$ ,  $\Sigma_{\setminus t} = \mathbf{Var}(\mathbf{y}_{\setminus t})$ ,  $\Sigma_{\setminus t, t} = \text{Cov}(\mathbf{y}_{\setminus t}, \mathbf{y}_t)$  and

$$\begin{aligned}
\Sigma_{t|\setminus t} &= \Sigma_t - \Sigma'_{\setminus t, t} \Sigma_{\setminus t}^{-1} \Sigma_{\setminus t, t} \\
&= \text{Var}(\mathbf{y}_t) - \text{Cov}(\mathbf{y}_t, \mathbf{y}_{\setminus t}) \text{Var}(\mathbf{y}_{\setminus t})^{-1} \text{Cov}(\mathbf{y}_{\setminus t}, \mathbf{y}_t) \\
&= \text{MSE}[\mathbf{L}(\mathbf{y}_t | \mathbf{y}_{\setminus t})] \\
&= \text{MSE}(\hat{\mathbf{y}}_t) \tag{3.18}
\end{aligned}$$

Thus,  $\mathbf{w}_t$ , the component corresponding to  $\mathbf{y}_t$ , is

$$\begin{aligned}
\mathbf{w}_t &= -\Sigma_{t|\setminus t}^{-1} \Sigma'_{\setminus t, t} \Sigma_{\setminus t}^{-1} [\mathbf{y}_{\setminus t} - \mathbf{E}(\mathbf{y}_{\setminus t})] + \Sigma_{t|\setminus t}^{-1} [\mathbf{y}_t - \mathbf{E}(\mathbf{y}_t)] \\
&= \text{MSE}(\hat{\mathbf{y}}_t)^{-1} \left[ \mathbf{y}_t - \mathbf{E}(\mathbf{y}_t) - \text{Cov}(\mathbf{y}_t, \mathbf{y}_{\setminus t}) \text{Var}(\mathbf{y}_{\setminus t})^{-1} \left\{ \mathbf{y}_{\setminus t} - \mathbf{E}(\mathbf{y}_{\setminus t}) \right\} \right] \\
&= \text{MSE}(\hat{\mathbf{y}}_t)^{-1} \left[ \mathbf{y}_t - \mathbf{L}(\mathbf{y}_t | \mathbf{y}_{\setminus t}) \right] \\
&= \text{MSE}(\hat{\mathbf{y}}_t)^{-1} (\mathbf{y}_t - \hat{\mathbf{y}}_t) \tag{3.19}
\end{aligned}$$

To proceed further, we need to evaluate  $\mathbf{S}$  and hence the block matrix  $\mathbf{A} \equiv \underline{\mathbf{G}} \mathbf{S} \underline{\mathbf{Z}}' + \underline{\mathbf{G}} \underline{\mathbf{G}}'$ . Each state  $\alpha_t$  only depends on errors up to time  $t-1$ , thus  $\text{Cov}(\varepsilon_{t+\tau}, \alpha_t) = 0$ , for  $\tau \geq 0$ . For error terms prior to time  $t$ , the covariance is

$$\begin{aligned}
\text{Cov}(\boldsymbol{\varepsilon}_{t-\tau}, \boldsymbol{\alpha}_t) &= \text{Cov}(\boldsymbol{\varepsilon}_{t-\tau}, \mathbf{T}_{t-1}\boldsymbol{\alpha}_{t-1} + \mathbf{H}_{t-1}\boldsymbol{\varepsilon}_{t-1}) \\
&= \text{Cov}(\boldsymbol{\varepsilon}_{t-\tau}, \boldsymbol{\alpha}_{t-1})\mathbf{T}'_{t-1} \\
&= \text{Cov}(\boldsymbol{\varepsilon}_{t-\tau}, \boldsymbol{\alpha}_{t-\tau+1})\mathbf{T}'_{t-\tau+1} \cdots \mathbf{T}'_{t-1} \\
&= \text{Cov}(\boldsymbol{\varepsilon}_{t-\tau}, \mathbf{T}_{t-\tau}\boldsymbol{\alpha}_{t-\tau} + \mathbf{H}_{t-\tau}\boldsymbol{\varepsilon}_{t-\tau})\underline{\mathbf{T}}'_{t-\tau+1, t-1} \\
&= \mathbf{H}'_{t-\tau}\underline{\mathbf{T}}'_{t-\tau+1, t-1} \tag{3.20}
\end{aligned}$$

where  $\tau > 0$ , and  $\underline{\mathbf{T}}_{t-\tau+1, t-1}$  is defined as

$$\underline{\mathbf{T}}_{t-\tau+1, t-1} = \begin{cases} \mathbf{T}_{t-1} \cdots \mathbf{T}_{t-\tau+1} & \tau = 3, 4, \dots \\ \mathbf{T}_{t-1} & \tau = 2 \\ \mathbf{I} & \tau = 1 \end{cases} \tag{3.21}$$

Thus, the  $(i, j)^{th}$  block of  $\mathbf{S}$  is  $\mathbf{H}'_i \underline{\mathbf{T}}'_{i+1, j-1}$  for  $i < j$  and zero otherwise. The matrices  $\underline{\mathbf{G}}$  and  $\underline{\mathbf{Z}}$  are both block-diagonal, hence the  $(i, j)^{th}$  block of  $\underline{\mathbf{G}}\mathbf{S}\underline{\mathbf{Z}}'$  is equal to  $\mathbf{G}_i \mathbf{H}'_i \underline{\mathbf{T}}'_{i+1, j-1} \mathbf{Z}'_j$  for  $i < j$  and zero otherwise. The matrix  $\underline{\mathbf{G}}\underline{\mathbf{G}}'$  is also block-diagonal. Grouping these results together, we can write

$$\mathbf{A} = \begin{bmatrix} \mathbf{G}_1 \mathbf{G}'_1 & \mathbf{G}_1 \mathbf{H}'_1 \mathbf{Z}'_2 & \mathbf{G}_1 \mathbf{H}'_1 \mathbf{T}'_2 \mathbf{Z}'_3 & \cdots & \mathbf{G}_1 \mathbf{H}'_1 \underline{\mathbf{T}}'_{2, n-1} \mathbf{Z}'_n \\ 0 & \mathbf{G}_2 \mathbf{G}'_2 & \mathbf{G}_2 \mathbf{H}'_2 \mathbf{Z}'_3 & \cdots & \mathbf{G}_2 \mathbf{H}'_2 \underline{\mathbf{T}}'_{3, n-1} \mathbf{Z}'_n \\ 0 & 0 & \mathbf{G}_3 \mathbf{G}'_3 & \cdots & \mathbf{G}_3 \mathbf{H}'_3 \underline{\mathbf{T}}'_{4, n-1} \mathbf{Z}'_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{G}_n \mathbf{G}'_n \end{bmatrix}$$

The matrix  $\mathbf{A}$  is invertible because it is block-upper-triangular and all the diagonal blocks are non-degenerate variance matrices, hence of full rank. If we let  $\mathbf{d}_t \equiv \mathbf{G}_t \mathbf{G}'_t \mathbf{u}_t + \mathbf{G}_t \mathbf{H}'_t \mathbf{r}_t$  and define  $\mathbf{w}$ ,  $\mathbf{d}$  as the stacked vectors of  $\mathbf{w}_t$ ,  $\mathbf{d}_t$  (respectively) for  $t = 1, \dots, n$ , we can write

$$\mathbf{d} = \mathbf{A} \mathbf{w} \quad (3.22)$$

We will now prove that  $\mathbf{w}_t = \mathbf{u}_t$  (from the KFS) is the unique solution to equation (3.22). Since  $\mathbf{A}$  is invertible, we can write  $\mathbf{w} = \mathbf{A}^{-1} \mathbf{d}$ , so the equation has a unique solution. Thus, it suffices to prove that  $\mathbf{w}_t = \mathbf{u}_t$  satisfies (3.22).

Repeated application of the KFS recursions yields

$$\begin{aligned} \mathbf{r}_t &= \mathbf{Z}'_{t+1} \mathbf{u}_{t+1} + \mathbf{T}'_{t+1} \mathbf{r}_{t+1} \\ &= \mathbf{Z}'_{t+1} \mathbf{u}_{t+1} + \mathbf{T}'_{t+1} (\mathbf{Z}'_{t+2} \mathbf{u}_{t+2} + \mathbf{T}'_{t+2} \mathbf{r}_{t+2}) \\ &= \mathbf{Z}'_{t+1} \mathbf{u}_{t+1} + \mathbf{T}'_{t+1} \mathbf{Z}'_{t+2} \mathbf{u}_{t+2} + \mathbf{T}'_{t+1} \mathbf{T}'_{t+2} (\mathbf{Z}'_{t+3} \mathbf{u}_{t+3} + \mathbf{T}'_{t+3} \mathbf{r}_{t+3}) \\ &= \mathbf{Z}'_{t+1} \mathbf{u}_{t+1} + \mathbf{T}'_{t+1} \mathbf{Z}'_{t+2} \mathbf{u}_{t+2} + \dots + \mathbf{T}'_{t+1, n-1} \mathbf{Z}'_n \mathbf{u}_n + \mathbf{T}'_{t+1, n} \mathbf{r}_n \\ &= \sum_{j=t+1}^n \mathbf{T}'_{t+1, j-1} \mathbf{Z}'_j \mathbf{u}_j \end{aligned}$$

because  $\mathbf{r}_n = 0$ . Thus, if we set  $\mathbf{w}_t = \mathbf{u}_t$ , the  $t^{th}$  vector component of the RHS of (3.22) is

$$\begin{aligned}
& \begin{bmatrix} 0 & \cdots & 0 & G_t G'_t & G_t H'_t Z'_{t+1} & \cdots & G_t H'_t \underline{T}'_{t+1, n-1} Z'_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} \\
&= G_t G'_t \mathbf{u}_t + G_t H'_t \sum_{j=t+1}^n \underline{T}'_{t+1, j-1} Z'_j \mathbf{u}_j \\
&= G_t G'_t \mathbf{u}_t + G_t H'_t \mathbf{r}_t \\
&= \mathbf{d}_t
\end{aligned} \tag{3.23}$$

as required. Completing the proof of Theorem 3.2.1 is now straightforward. We can write

$$\begin{aligned}
\mathbf{u}_t &= \text{MSE}(\hat{\mathbf{y}}_t)^{-1} (\mathbf{y}_t - \hat{\mathbf{y}}_t) \\
\Rightarrow \hat{\mathbf{y}}_t &= \mathbf{y}_t - \text{MSE}(\hat{\mathbf{y}}_t) \mathbf{u}_t
\end{aligned} \tag{3.24}$$

Thus, the MSE of  $\hat{\mathbf{y}}_t$  as an estimator of  $\mathbf{y}_t$  is

$$\begin{aligned}
\text{MSE}(\hat{\mathbf{y}}_t) &= \text{MSE}\{\mathbf{y}_t - \text{MSE}(\hat{\mathbf{y}}_t) \mathbf{u}_t\} \\
&= \text{Var}\{\text{MSE}(\hat{\mathbf{y}}_t) \mathbf{u}_t\} \\
&= \text{MSE}(\hat{\mathbf{y}}_t) \text{Var}(\mathbf{u}_t) \text{MSE}(\hat{\mathbf{y}}_t) \\
\Rightarrow \text{MSE}(\hat{\mathbf{y}}_t) &= \text{Var}(\mathbf{u}_t)^{-1}
\end{aligned} \tag{3.25}$$

In the technical introduction we established that  $\mathbf{M}_t = \text{Var}(\mathbf{u}_t)$ . Substituting into (3.24) and (3.25) yields



$$\hat{\mathbf{y}}_t = \mathbf{y}_t - \mathbf{M}_t^{-1} \mathbf{u}_t \quad (3.26)$$

and

$$\text{MSE}(\hat{\mathbf{y}}_t) = \mathbf{M}_t^{-1} \quad (3.27)$$

as required.

□

### 3.2.3 State Estimate

We will now establish a formula for the interpolated MMSLE of the state vector. We will rely on the following simple corollary of (2.5.2):

**Corollary 3.2.2.** *Let  $\mathbf{a}, \mathbf{b}$  be random vectors. If  $g(\mathbf{b})$  is a linear function of  $\mathbf{b}$ , then*

$$\text{Cov}[g(\mathbf{b}), \mathbf{a} - \mathbf{L}(\mathbf{a}|\mathbf{b})] = 0 \quad (3.28)$$

*Proof.* The function  $g(\mathbf{b})$  is linear, thus we can write  $g(\mathbf{b}) = \mathbf{C}\mathbf{b} + \mathbf{k}$ , where  $\mathbf{C}, \mathbf{k}$  are constants. We have

$$\begin{aligned} \text{Cov}[g(\mathbf{b}), \mathbf{a} - \mathbf{L}(\mathbf{a}|\mathbf{b})] &= \text{Cov}[\mathbf{C}\mathbf{b} + \mathbf{k}, \mathbf{a} - \mathbf{L}(\mathbf{a}|\mathbf{b})] \\ &= \mathbf{C} \text{Cov}[\mathbf{b}, \mathbf{a} - \mathbf{L}(\mathbf{a}|\mathbf{b})] \\ &= 0 \end{aligned} \quad (3.29)$$

because the linear estimation error  $\mathbf{a} - \mathbf{L}(\mathbf{a}|\mathbf{b})$  is uncorrelated with  $\mathbf{b}$ .

□

The intuitive interpretation of this corollary is: the linear estimation error must be uncorrelated with any linear function of the known variable

**b.** If this were not the case, the function  $g(\mathbf{b})$  could be used to predict the error, thus it could be incorporated into  $L(\mathbf{a}|\mathbf{b})$  to produce a linear estimator with smaller MSE. However, the linear expectation  $L(\mathbf{a}|\mathbf{b})$  minimises the MSE within the class of all linear estimators of  $\mathbf{a}$  given  $\mathbf{b}$ , so this cannot be possible.

If  $\mathbf{a}, \mathbf{b}$  are jointly normally distributed, Corollary 3.2.2 can be extended to any function  $g(\mathbf{b})$ , as long as it is uniquely defined by  $\mathbf{b}$ . By a similar argument, the estimation error  $\mathbf{a} - E(\mathbf{a}|\mathbf{b})$  is uncorrelated with  $g(\mathbf{b})$ , since it is not possible to improve on the MMSE estimator  $E(\mathbf{a}|\mathbf{b})$ .

**Theorem 3.2.3.** *The interpolated MMSLE of state  $\alpha_t$  is given by*

$$\hat{\alpha}_t = \hat{\alpha}_t - \mathbf{P}_t \mathbf{R}_t \mathbf{M}_t^{-1} \mathbf{u}_t \quad (3.30)$$

and its mean square error as an estimator of  $\alpha_t$  is

$$\text{MSE}(\hat{\alpha}_t) = \mathbf{P}_t + \mathbf{P}_t \mathbf{R}_t \mathbf{M}_t^{-1} \mathbf{R}_t' \mathbf{P}_t \quad (3.31)$$

where  $\mathbf{R}_t \equiv \mathbf{Z}_t' \mathbf{F}_t^{-1} - \mathbf{L}_t' \mathbf{N}_t \mathbf{K}_t$ .

*Proof.* Let  $\mathbf{a} = \mathbf{y}_t$  and  $\mathbf{b} = \mathbf{Y}_n^{(t)}$ , so that  $L(\mathbf{a}|\mathbf{b}) = \hat{\mathbf{y}}_t$ . Setting  $g(\mathbf{b}) = \mathbf{y}_j$ , for  $j \neq t$ , satisfies the conditions of Corollary 3.2.2 and we can deduce:

$$\text{Cov}(\mathbf{y}_t - \hat{\mathbf{y}}_t, \mathbf{y}_j) = 0$$

Furthermore,  $\hat{\mathbf{y}}_t$  is a linear combination of the elements of  $\mathbf{Y}_n^{(t)}$ , hence the span of  $\{\mathbf{Y}_n^{(t)}, \mathbf{y}_t\}$  is the same as the span of  $\{\mathbf{Y}_n^{(t)}, \mathbf{y}_t - \hat{\mathbf{y}}_t\}$ . We can make use of this fact to write

$$\begin{aligned}
\hat{\alpha}_t &= L(\alpha_t | Y_n^{(t)}, \mathbf{y}_t) \\
&= L(\alpha_t | Y_n^{(t)}, \mathbf{y}_t - \dot{\mathbf{y}}_t) \\
&= L(\alpha_t | Y_n^{(t)}) + \text{Cov}(\alpha_t, \mathbf{y}_t - \dot{\mathbf{y}}_t) \text{Var}(\mathbf{y}_t - \dot{\mathbf{y}}_t)^{-1} (\mathbf{y}_t - \dot{\mathbf{y}}_t) \\
&= \hat{\alpha}_t + \text{Cov}(\alpha_t, \mathbf{y}_t - \dot{\mathbf{y}}_t) \text{MSE}(\dot{\mathbf{y}}_t)^{-1} (\mathbf{y}_t - \dot{\mathbf{y}}_t) \\
&= \hat{\alpha}_t + \text{Cov}(\alpha_t, \mathbf{y}_t - \dot{\mathbf{y}}_t) \mathbf{u}_t \\
&\Leftrightarrow \hat{\alpha}_t = \hat{\alpha}_t - \text{Cov}(\alpha_t, \mathbf{y}_t - \dot{\mathbf{y}}_t) \mathbf{u}_t \tag{3.32}
\end{aligned}$$

The covariance term in this expression is

$$\begin{aligned}
\text{Cov}(\alpha_t, \mathbf{y}_t - \dot{\mathbf{y}}_t) &= \text{Cov}(\alpha_t, \mathbf{M}_t^{-1} \mathbf{u}_t) \\
&= \text{Cov}(\alpha_t, \mathbf{u}_t) \mathbf{M}_t^{-1} \\
&= \text{Cov}(\alpha_t, \mathbf{F}_t^{-1} \mathbf{v}_t - \mathbf{K}_t' \mathbf{r}_t) \mathbf{M}_t^{-1} \\
&= \{\text{Cov}(\alpha_t, \mathbf{v}_t) \mathbf{F}_t^{-1} - \text{Cov}(\alpha_t, \mathbf{r}_t) \mathbf{K}_t\} \mathbf{M}_t^{-1} \tag{3.33}
\end{aligned}$$

The remaining terms are straightforward to evaluate:

$$\begin{aligned}
\text{Cov}(\alpha_t, \mathbf{v}_t) &= \mathbf{P}_t \underline{\mathbf{L}}_{t,t-1}' \mathbf{Z}_t' \\
&= \mathbf{P}_t \mathbf{Z}_t' \tag{3.34}
\end{aligned}$$

by definition of  $\underline{\mathbf{L}}_{s,t}$ . For the final term, we apply Corollary 3.2.2 to  $\mathbf{a} = \alpha_t$ ,  $\mathbf{b} = \mathbf{Y}_n$ , and  $g(\mathbf{b}) = \mathbf{r}_t$ . In this case  $L(\mathbf{a} | \mathbf{b}) = \hat{\alpha}_t$ , so we can

write  $\text{Cov}(\boldsymbol{\alpha}_t - \hat{\boldsymbol{\alpha}}_t, \mathbf{r}_t) = 0$ . The vector  $\mathbf{r}_t$  is a weighted sum of future innovations  $\mathbf{v}_{t+1}, \dots, \mathbf{v}_n$  hence is uncorrelated to the estimator  $\mathbf{a}_t$ , which (as can be seen in the KF recursions) is a weighted sum of innovations  $\mathbf{v}_1, \dots, \mathbf{v}_t$ . We have

$$\begin{aligned}
\text{Cov}(\boldsymbol{\alpha}_t, \mathbf{r}_t) &= \text{Cov}(\boldsymbol{\alpha}_t - \hat{\boldsymbol{\alpha}}_t + \hat{\boldsymbol{\alpha}}_t, \mathbf{r}_t) \\
&= \text{Cov}(\boldsymbol{\alpha}_t - \hat{\boldsymbol{\alpha}}_t, \mathbf{r}_t) + \text{Cov}(\hat{\boldsymbol{\alpha}}_t, \mathbf{r}_t) \\
&= \text{Cov}(\mathbf{a}_t + \mathbf{P}_t \mathbf{r}_{t-1}, \mathbf{r}_t) \\
&= \text{Cov}(\mathbf{a}_t, \mathbf{r}_t) + \mathbf{P}_t \text{Cov}(\mathbf{r}_{t-1}, \mathbf{r}_t) \\
&= \mathbf{P}_t \text{Cov}(\mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{L}_t' \mathbf{r}_t, \mathbf{r}_t) \\
&= \mathbf{P}_t \mathbf{Z}_t' \mathbf{F}_t^{-1} \text{Cov}(\mathbf{v}_t, \mathbf{r}_t) + \mathbf{P}_t \mathbf{L}_t' \text{Var}(\mathbf{r}_t) \\
&= \mathbf{P}_t \mathbf{L}_t' \mathbf{N}_t
\end{aligned} \tag{3.35}$$

Substituting expressions (3.33), (3.34) and (3.35) into (3.32) yields

$$\begin{aligned}
\dot{\boldsymbol{\alpha}}_t &= \hat{\boldsymbol{\alpha}}_t - (\mathbf{P}_t \mathbf{Z}_t' \mathbf{F}_t^{-1} - \mathbf{P}_t \mathbf{L}_t' \mathbf{N}_t \mathbf{K}_t) \mathbf{M}_t^{-1} \mathbf{u}_t \\
&= \hat{\boldsymbol{\alpha}}_t - \mathbf{P}_t (\mathbf{Z}_t' \mathbf{F}_t^{-1} - \mathbf{L}_t' \mathbf{N}_t \mathbf{K}_t) \mathbf{M}_t^{-1} \mathbf{u}_t
\end{aligned} \tag{3.36}$$

and the expression inside the brackets is equal to  $\mathbf{R}_t$ , by definition.

□

### 3.3 Partially Interpolated Estimates

#### 3.3.1 Introduction

Assume that we wish to find the interpolated estimate of  $\mathbf{y}_t$ , but we only want to take into account the data observed by time  $m$ , where  $t < m \leq n$ . We define  $\hat{\mathbf{y}}_{t|m} \equiv E(\mathbf{y}_t | \mathbf{Y}_m^{(t)})$  and  $\hat{\boldsymbol{\alpha}}_{t|m} \equiv E(\boldsymbol{\alpha}_t | \mathbf{Y}_m^{(t)})$ . Note that  $\hat{\mathbf{y}}_{t|n} = \hat{\mathbf{y}}_t$  and  $\hat{\boldsymbol{\alpha}}_{t|n} = \hat{\boldsymbol{\alpha}}_t$ . We will first prove the following lemma:

**Lemma 3.3.1.**

$$\underline{\mathbf{T}}'_{t+1,m} - \sum_{j=t+1}^m \underline{\mathbf{T}}'_{t+1,j-1} \mathbf{Z}'_j \mathbf{K}'_j \underline{\mathbf{L}}'_{j+1,m} = \underline{\mathbf{L}}'_{t+1,m} \quad (3.37)$$

for  $0 < t < m \leq n$

*Proof.* By backward induction on  $t$ . For  $t = m-1$  the expression reduces to

$$\mathbf{T}'_m - \mathbf{Z}'_m \mathbf{K}'_m = \mathbf{L}'_m \quad (3.38)$$

which is true by definition of  $\mathbf{L}_t$ . We now assume the result holds for  $t = \tau > 1$  and consider the expression for  $t = \tau - 1$

$$\begin{aligned}
& \underline{\mathbf{T}}'_{\tau,m} - \sum_{j=\tau}^m \underline{\mathbf{T}}'_{\tau,j-1} \mathbf{Z}'_j \mathbf{K}'_j \underline{\mathbf{L}}'_{j+1,m} \\
&= \underline{\mathbf{T}}'_{\tau,m} - \sum_{j=\tau+1}^m \underline{\mathbf{T}}'_{\tau,j-1} \mathbf{Z}'_j \mathbf{K}'_j \underline{\mathbf{L}}'_{j+1,m} - \mathbf{Z}'_{\tau} \mathbf{K}'_{\tau} \underline{\mathbf{L}}'_{\tau+1,m} \\
&= \mathbf{T}'_{\tau} \left( \underline{\mathbf{T}}'_{\tau+1,m} - \sum_{j=\tau+1}^m \underline{\mathbf{T}}'_{\tau+1,j-1} \mathbf{Z}'_j \mathbf{K}'_j \underline{\mathbf{L}}'_{j+1,m} \right) - \mathbf{Z}'_{\tau} \mathbf{K}'_{\tau} \underline{\mathbf{L}}'_{\tau+1,m} \\
&= \mathbf{T}'_{\tau} \underline{\mathbf{L}}'_{\tau+1,m} - \mathbf{Z}'_{\tau} \mathbf{K}'_{\tau} \underline{\mathbf{L}}'_{\tau+1,m} \\
&= (\mathbf{T}'_{\tau} - \mathbf{Z}'_{\tau} \mathbf{K}'_{\tau}) \underline{\mathbf{L}}'_{\tau+1,m} \\
&= \mathbf{L}'_{\tau} \underline{\mathbf{L}}'_{\tau+1,m} \\
&= \underline{\mathbf{L}}'_{\tau,m}
\end{aligned}$$

□

### 3.3.2 Observation Estimate

**Theorem 3.3.2.** *The interpolated estimate of  $\mathbf{y}_t$  conditional on the punctured space  $\mathbf{Y}_m^{(t)}$  is given by*

$$\hat{\mathbf{y}}_{t|m} = \mathbf{y}_t - \mathbf{M}_{t|m}^{-1} \mathbf{u}_{t|m} \quad (3.39)$$

and its mean square error is  $\mathbf{M}_{t|m}^{-1}$ , where

$$\mathbf{u}_{t|m} \equiv \mathbf{u}_t + \mathbf{K}'_t \underline{\mathbf{L}}'_{t+1,m} \mathbf{r}_m \quad (3.40)$$

$$\mathbf{M}_{t|m} \equiv \mathbf{M}_t - \mathbf{K}'_t \underline{\mathbf{L}}'_{t+1,m} \mathbf{N}_m \underline{\mathbf{L}}_{t+1,m} \mathbf{K}_t \quad (3.41)$$

*Proof.* Define  $\hat{\mathbf{x}}_{t|m} \equiv \mathbf{L}(\mathbf{x}_t|\mathbf{Y}_m)$  and  $\hat{\boldsymbol{\alpha}}_{t|m} \equiv \mathbf{L}(\boldsymbol{\alpha}_t|\mathbf{Y}_m)$ , the smoothed estimate of the signal and state vectors conditional on data up to time  $m$ . We have

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_t &= \mathbf{L}(\boldsymbol{\alpha}_t|\mathbf{Y}_n) \\
&= \mathbf{L}(\boldsymbol{\alpha}_t|\mathbf{Y}_m, \mathbf{v}_{m+1}, \dots, \mathbf{v}_n) \\
&= \mathbf{L}(\boldsymbol{\alpha}_t|\mathbf{Y}_m) + \sum_{s=m+1}^n \text{Cov}(\boldsymbol{\alpha}_t, \mathbf{v}_s) \mathbf{F}_s^{-1} \mathbf{v}_s \\
&= \hat{\boldsymbol{\alpha}}_{t|m} + \sum_{s=m+1}^n \mathbf{P}_t \underline{\mathbf{L}}'_{t,s-1} \mathbf{Z}'_s \mathbf{F}_s^{-1} \mathbf{v}_s \\
&= \hat{\boldsymbol{\alpha}}_{t|m} + \mathbf{P}_t \underline{\mathbf{L}}'_{t,m} \mathbf{r}_m
\end{aligned} \tag{3.42}$$

thus

$$\begin{aligned}
\hat{\mathbf{x}}_{t|m} &= \mathbf{L}(\mathbf{x}_t|\mathbf{Y}_m) \\
&= \mathbf{Z}_t \mathbf{L}(\boldsymbol{\alpha}_t|\mathbf{Y}_m) \\
&= \mathbf{Z}_t (\hat{\boldsymbol{\alpha}}_t - \mathbf{P}_t \underline{\mathbf{L}}'_{t,m} \mathbf{r}_m) \\
&= \hat{\mathbf{x}}_t - \mathbf{Z}_t \mathbf{P}_t \underline{\mathbf{L}}'_{t,m} \mathbf{r}_m
\end{aligned} \tag{3.43}$$

We can use this result to write

$$\begin{aligned}
\mathbf{y}_t - \hat{\mathbf{x}}_{t|m} &= \mathbf{y}_t - \hat{\mathbf{x}}_t + \mathbf{Z}_t \mathbf{P}_t \underline{\mathbf{L}}'_{t,m} \mathbf{r}_m \\
&= \mathbf{G}_t \mathbf{G}'_t \mathbf{u}_t + \mathbf{G}_t \mathbf{H}'_t \mathbf{r}_t + \mathbf{Z}_t \mathbf{P}_t \underline{\mathbf{L}}'_{t,m} \mathbf{r}_m
\end{aligned} \tag{3.44}$$

where the expression for  $\mathbf{y}_t - \hat{\mathbf{x}}_t$  is from the proof of Theorem 3.2.1. We now define stacked vectors in a similar way. Suppose that  $\mathbf{y}_{[m]}$ ,  $\boldsymbol{\alpha}_{[m]}$ ,  $\boldsymbol{\varepsilon}_{[m]}$ ,  $\mathbf{x}_{[m]}$  and  $\hat{\mathbf{x}}_{[m]}$  are (respectively) the stacked vectors of  $\mathbf{y}_t$ ,  $\boldsymbol{\alpha}_t$ ,  $\boldsymbol{\varepsilon}_t$ ,  $\mathbf{x}_t$  and  $\hat{\mathbf{x}}_{t|m}$  for  $t = 1, \dots, m$ . We also define the block-diagonal matrices  $\underline{\mathbf{G}}_{[m]} \equiv \text{diag}\{\mathbf{G}_1, \dots, \mathbf{G}_m\}$  and  $\underline{\mathbf{Z}}_{[m]} \equiv \text{diag}\{\mathbf{Z}_1, \dots, \mathbf{Z}_m\}$ . As before, we can write the measurement equation in stack vector form as  $\mathbf{y}_{[m]} = \mathbf{x}_{[m]} + \underline{\mathbf{G}}_{[m]} \boldsymbol{\varepsilon}_{[m]} = \underline{\mathbf{Z}}_{[m]} \boldsymbol{\alpha}_{[m]} + \underline{\mathbf{G}}_{[m]} \boldsymbol{\varepsilon}_{[m]}$ . Using the linear statistics results from the technical introduction, we have

$$\begin{aligned}
\hat{\mathbf{x}}_{[m]} &= \mathbf{L}(\mathbf{x}_{[m]}|\mathbf{y}_{[m]}) \\
&= \mathbf{E}(\mathbf{x}_{[m]}) + \text{Cov}(\mathbf{x}_{[m]}, \mathbf{y}_{[m]})\text{Var}(\mathbf{y}_{[m]})^{-1}\{\mathbf{y}_{[m]} - \mathbf{E}(\mathbf{y}_{[m]})\} \quad (3.45)
\end{aligned}$$

As before,  $\mathbf{E}(\boldsymbol{\varepsilon}_{[m]}) = 0$ , thus  $\mathbf{E}(\mathbf{x}_{[m]}) = \mathbf{E}(\mathbf{y}_{[m]})$ . We can write the covariance matrix of  $\mathbf{x}_{[m]}$  and  $\mathbf{y}_{[m]}$  as

$$\begin{aligned}
\text{Cov}(\mathbf{x}_{[m]}, \mathbf{y}_{[m]}) &= \text{Cov}(\mathbf{y}_{[m]} - \underline{\mathbf{G}}_{[m]}\boldsymbol{\varepsilon}_{[m]}, \mathbf{y}_{[m]}) \\
&= \text{Var}(\mathbf{y}_{[m]}) - \underline{\mathbf{G}}_{[m]}\text{Cov}(\boldsymbol{\varepsilon}_{[m]}, \mathbf{y}_{[m]}) \\
&= \text{Var}(\mathbf{y}_{[m]}) - \underline{\mathbf{G}}_{[m]}\text{Cov}(\boldsymbol{\varepsilon}_{[m]}, \underline{\mathbf{Z}}_{[m]}\boldsymbol{\alpha}_{[m]} + \underline{\mathbf{G}}_{[m]}\boldsymbol{\varepsilon}_{[m]}) \\
&= \text{Var}(\mathbf{y}_{[m]}) - \underline{\mathbf{G}}_{[m]}\text{Cov}(\boldsymbol{\varepsilon}_{[m]}, \boldsymbol{\alpha}_{[m]})\underline{\mathbf{Z}}'_{[m]} \\
&\quad - \underline{\mathbf{G}}_{[m]}\text{Var}(\boldsymbol{\varepsilon}_{[m]})\underline{\mathbf{G}}'_{[m]} \\
&= \boldsymbol{\Sigma}_{[m]} - \underline{\mathbf{G}}_{[m]}\mathbf{S}_{[m]}\underline{\mathbf{Z}}'_{[m]} - \underline{\mathbf{G}}_{[m]}\underline{\mathbf{G}}'_{[m]} \quad (3.46)
\end{aligned}$$

where  $\boldsymbol{\Sigma}_{[m]} \equiv \text{Var}(\mathbf{y}_{[m]})$  and  $\mathbf{S}_{[m]} \equiv \text{Cov}(\boldsymbol{\varepsilon}_{[m]}, \boldsymbol{\alpha}_{[m]})$ . Substituting back into (3.45), we have

$$\begin{aligned}
\hat{\mathbf{x}}_{[m]} &= \mathbf{E}(\mathbf{y}_{[m]}) + (\boldsymbol{\Sigma}_{[m]} - \underline{\mathbf{G}}_{[m]}\mathbf{S}_{[m]}\underline{\mathbf{Z}}'_{[m]} - \underline{\mathbf{G}}_{[m]}\underline{\mathbf{G}}'_{[m]})\boldsymbol{\Sigma}_{[m]}^{-1}\{\mathbf{y}_{[m]} - \mathbf{E}(\mathbf{y}_{[m]})\} \\
&= \mathbf{E}(\mathbf{y}_{[m]}) + \{\mathbf{y}_{[m]} - \mathbf{E}(\mathbf{y}_{[m]})\} \\
&\quad - (\underline{\mathbf{G}}_{[m]}\mathbf{S}_{[m]}\underline{\mathbf{Z}}'_{[m]} + \underline{\mathbf{G}}_{[m]}\underline{\mathbf{G}}'_{[m]})\boldsymbol{\Sigma}_{[m]}^{-1}\{\mathbf{y}_{[m]} - \mathbf{E}(\mathbf{y}_{[m]})\} \\
&= \mathbf{y}_{[m]} - (\underline{\mathbf{G}}_{[m]}\mathbf{S}_{[m]}\underline{\mathbf{Z}}'_{[m]} + \underline{\mathbf{G}}_{[m]}\underline{\mathbf{G}}'_{[m]})\boldsymbol{\Sigma}_{[m]}^{-1}\{\mathbf{y}_{[m]} - \mathbf{E}(\mathbf{y}_{[m]})\} \quad (3.47)
\end{aligned}$$

which can be rearranged to give



$$\mathbf{y}_{[m]} - \hat{\mathbf{x}}_{[m]} = (\underline{\mathbf{G}}_{[m]} \mathbf{S}_{[m]} \underline{\mathbf{Z}}' + \underline{\mathbf{G}}_{[m]} \underline{\mathbf{G}}'_{[m]}) \underline{\Sigma}_{[m]}^{-1} \{\mathbf{y}_{[m]} - \mathbb{E}(\mathbf{y}_{[m]})\} \quad (3.48)$$

The LHS of (3.44) is the  $t^{th}$  vector component of the LHS of (3.48), thus the same must hold for the RHS. Let  $\mathbf{w}_{t|m}$  be the  $t^{th}$  vector component of  $\underline{\Sigma}_{[m]}^{-1} \{\mathbf{y}_{[m]} - \mathbb{E}(\mathbf{y}_{[m]})\}$ . From the previous proofs, we know that it is equal to

$$\mathbf{w}_{t|m} = \text{MSE}(\hat{\mathbf{y}}_{t|m})^{-1} (\mathbf{y}_t - \hat{\mathbf{y}}_{t|m}) \quad (3.49)$$

The matrix  $\mathbf{A}_{[m]} \equiv \underline{\mathbf{G}}_{[m]} \mathbf{S}_{[m]} \underline{\mathbf{Z}}' + \underline{\mathbf{G}}_{[m]} \underline{\mathbf{G}}'_{[m]}$  has the form

$$\mathbf{A}_{[m]} = \begin{bmatrix} \mathbf{G}_1 \mathbf{G}'_1 & \mathbf{G}_1 \mathbf{H}'_1 \mathbf{Z}'_2 & \mathbf{G}_1 \mathbf{H}'_1 \mathbf{T}'_2 \mathbf{Z}'_3 & \cdots & \mathbf{G}_1 \mathbf{H}'_1 \mathbf{T}'_{2,m-1} \mathbf{Z}'_m \\ 0 & \mathbf{G}_2 \mathbf{G}'_2 & \mathbf{G}_2 \mathbf{H}'_2 \mathbf{Z}'_3 & \cdots & \mathbf{G}_2 \mathbf{H}'_2 \mathbf{T}'_{3,m-1} \mathbf{Z}'_m \\ 0 & 0 & \mathbf{G}_3 \mathbf{G}'_3 & \cdots & \mathbf{G}_3 \mathbf{H}'_3 \mathbf{T}'_{4,m-1} \mathbf{Z}'_m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{G}_m \mathbf{G}'_m \end{bmatrix}$$

The matrix  $\mathbf{A}_{[m]}$  is block-upper-triangular and the diagonal blocks are all invertible matrices, hence  $\mathbf{A}_{[m]}$  is invertible. If we let  $\mathbf{d}_{t|m} \equiv \mathbf{G}_t \mathbf{G}'_t \mathbf{u}_t + \mathbf{G}_t \mathbf{H}'_t \mathbf{r}_t + \mathbf{Z}_t \mathbf{P}_t \underline{\mathbf{L}}'_{t,m} \mathbf{r}_m$  and define  $\mathbf{w}_{[m]}$ ,  $\mathbf{d}_{[m]}$  as the stacked vectors of  $\mathbf{w}_{t|m}$ ,  $\mathbf{d}_{t|m}$  (respectively) for  $t = 1, \dots, m$ , we can write

$$\mathbf{d}_{[m]} = \mathbf{A}_{[m]} \mathbf{w}_{[m]} \quad (3.50)$$

We will now proceed to prove that  $\mathbf{w}_{t|m} = \mathbf{u}_{t|m}$  is the unique solution to equation (3.50). Since  $\mathbf{A}_{[m]}$  is invertible, we can write  $\mathbf{w}_{[m]} = \mathbf{A}_{[m]}^{-1} \mathbf{d}_{[m]}$ ,

so the equation has a unique solution. It now suffices to prove that  $\mathbf{w}_{t|m} = \mathbf{u}_{t|m}$  satisfies (3.50).

It needs to be shown that the  $t^{th}$  vector component of  $\mathbf{A}_{[m]}\mathbf{u}_{[m]}$  is equal to  $\mathbf{v}_{t|m}$ . We have

$$\begin{aligned}
\{\mathbf{A}_{[m]}\mathbf{u}_{[m]}\}_t &= \mathbf{G}_t\mathbf{G}'_t\mathbf{u}_{t|m} + \mathbf{G}_t\mathbf{H}'_t \sum_{j=t+1}^m \mathbf{T}'_{t+1,j-1}\mathbf{Z}'_j\mathbf{u}_{j|m} \\
&= \mathbf{G}_t\mathbf{G}'_t(\mathbf{u}_t + \mathbf{K}'_t\mathbf{L}'_{t+1,m}\mathbf{r}_m) \\
&\quad + \mathbf{G}_t\mathbf{H}'_t \sum_{j=t+1}^m \mathbf{T}'_{t+1,j-1}\mathbf{Z}'_j(\mathbf{u}_j + \mathbf{K}'_j\mathbf{L}'_{j+1,m}\mathbf{r}_m) \\
&= \mathbf{G}_t\mathbf{G}'_t\mathbf{u}_t + \mathbf{G}_t\mathbf{H}'_t \sum_{j=t+1}^m \mathbf{T}'_{t+1,j-1}\mathbf{Z}'_j\mathbf{u}_j + \mathbf{G}_t\mathbf{G}'_t\mathbf{K}'_t\mathbf{L}'_{t+1,m}\mathbf{r}_m \\
&\quad + \mathbf{G}_t\mathbf{H}'_t \sum_{j=t+1}^m \mathbf{T}'_{t+1,j-1}\mathbf{Z}'_j\mathbf{K}'_j\mathbf{L}'_{j+1,m}\mathbf{r}_m \\
&= \mathbf{G}_t\mathbf{G}'_t\mathbf{u}_t + \mathbf{G}_t\mathbf{H}'_t \sum_{j=t+1}^m \mathbf{T}'_{t+1,j-1}\mathbf{Z}'_j\mathbf{u}_j + \mathbf{G}_t\mathbf{G}'_t\mathbf{K}'_t\mathbf{L}'_{t+1,m}\mathbf{r}_m \\
&\quad + \mathbf{G}_t\mathbf{H}'_t(\mathbf{T}'_{t+1,m} - \mathbf{L}'_{t+1,m})\mathbf{r}_m \tag{3.51}
\end{aligned}$$

using Lemma 3.37. The sum in this expression can be evaluated by noting that

$$\begin{aligned}
\mathbf{r}_t &= \mathbf{Z}'_{t+1} \mathbf{u}_{t+1} + \mathbf{T}'_{t+1} \mathbf{r}_{t+1} \\
&= \mathbf{Z}'_{t+1} \mathbf{u}_{t+1} + \mathbf{T}'_{t+1} (\mathbf{Z}'_{t+2} \mathbf{u}_{t+2} + \mathbf{T}'_{t+2} \mathbf{r}_{t+2}) \\
&= \mathbf{Z}'_{t+1} \mathbf{u}_{t+1} + \mathbf{T}'_{t+1} \mathbf{Z}'_{t+2} \mathbf{u}_{t+2} + \mathbf{T}'_{t+1} \mathbf{T}'_{t+2} (\mathbf{Z}'_{t+3} \mathbf{u}_{t+3} + \mathbf{T}'_{t+3} \mathbf{r}_{t+3}) \\
&= \mathbf{Z}'_{t+1} \mathbf{u}_{t+1} + \mathbf{T}'_{t+1} \mathbf{Z}'_{t+2} \mathbf{u}_{t+2} + \dots + \mathbf{T}'_{t+1, m-1} \mathbf{Z}'_m \mathbf{u}_m + \mathbf{T}'_{t+1, m} \mathbf{r}_m \\
&\Rightarrow \sum_{j=t+1}^m \mathbf{T}'_{t+1, j-1} \mathbf{Z}'_j \mathbf{u}_j = \mathbf{r}_t - \mathbf{T}'_{t+1, m} \mathbf{r}_m \tag{3.52}
\end{aligned}$$

Substituting into (3.51) yields

$$\begin{aligned}
\{\mathbf{A}_{[m]} \mathbf{u}_{[m]}\}_t &= \mathbf{G}_t \mathbf{G}'_t \mathbf{u}_t + \mathbf{G}_t \mathbf{H}'_t (\mathbf{r}_t - \mathbf{T}'_{t+1, m} \mathbf{r}_m) + \mathbf{G}_t \mathbf{G}'_t \mathbf{K}'_t \mathbf{L}'_{t+1, m} \mathbf{r}_m \\
&\quad + \mathbf{G}_t \mathbf{H}'_t (\mathbf{T}'_{t+1, m} - \mathbf{L}'_{t+1, m}) \mathbf{r}_m \\
&= \mathbf{G}_t \mathbf{G}'_t \mathbf{u}_t + \mathbf{G}_t \mathbf{H}'_t \mathbf{r}_t + \mathbf{G}_t \mathbf{G}'_t \mathbf{K}'_t \mathbf{L}'_{t+1, m} \mathbf{r}_m \\
&\quad - \mathbf{G}_t \mathbf{H}'_t \mathbf{L}'_{t+1, m} \mathbf{r}_m \\
&= \mathbf{G}_t \mathbf{G}'_t \mathbf{u}_t + \mathbf{G}_t \mathbf{H}'_t \mathbf{r}_t + (\mathbf{G}_t \mathbf{G}'_t \mathbf{K}'_t - \mathbf{G}_t \mathbf{H}'_t) \mathbf{L}'_{t+1, m} \mathbf{r}_m \tag{3.53}
\end{aligned}$$

The expression in the brackets can be simplified using the Kalman filter equations. We have

$$\begin{aligned}
G_t G_t' K_t' - G_t H_t' &= G_t G_t' K_t' - (F_t K_t' - Z_t P_t T_t') \\
&= (G_t G_t' - F_t) K_t' - Z_t P_t T_t' \\
&= Z_t P_t Z_t' K_t' - Z_t P_t T_t' \\
&= Z_t P_t L_t'
\end{aligned} \tag{3.54}$$

Hence

$$\{\mathbf{A}_{[m]} \mathbf{u}_{[m]}\}_t = \mathbf{G}_t \mathbf{G}_t' \mathbf{u}_t + \mathbf{G}_t \mathbf{H}_t' \mathbf{r}_t + \mathbf{Z}_t \mathbf{P}_t \mathbf{L}_t' \underline{\mathbf{L}}_{t+1,m}' \mathbf{r}_m = \mathbf{d}_{t|m} \tag{3.55}$$

Thus,  $\mathbf{u}_{[m]}$  satisfies equation (3.50) and we can write

$$\begin{aligned}
\mathbf{u}_{t|m} &= \text{MSE}(\hat{\mathbf{y}}_{t|m})_{\mathbf{y}_t}^{-1} (\mathbf{y}_t - \hat{\mathbf{y}}_{t|m}) \\
\Rightarrow \hat{\mathbf{y}}_{t|m} &= \mathbf{y}_t - \text{MSE}(\hat{\mathbf{y}}_{t|m})_{\mathbf{y}_t} \mathbf{u}_{t|m}
\end{aligned} \tag{3.56}$$

To complete the proof, we need to compute the MSE of  $\hat{\mathbf{y}}_{t|m}$ :

$$\begin{aligned}
\text{MSE}(\hat{\mathbf{y}}_{t|m})_{\mathbf{y}_t} &= \text{MSE}_{\mathbf{y}_t} \left( \mathbf{y}_t - \text{MSE}(\hat{\mathbf{y}}_{t|m})_{\mathbf{y}_t} \mathbf{u}_{t|m} \right) \\
&= \text{Cov} \left( \text{MSE}(\hat{\mathbf{y}}_{t|m})_{\mathbf{y}_t} \mathbf{u}_{t|m} \right) \\
&= \text{MSE}(\hat{\mathbf{y}}_{t|m})_{\mathbf{y}_t} \text{Var}(\mathbf{u}_{t|m}) \text{MSE}(\hat{\mathbf{y}}_{t|m})_{\mathbf{y}_t}
\end{aligned} \tag{3.57}$$

$$\begin{aligned}
\Rightarrow \text{MSE}(\dot{\mathbf{y}}_{t|m}) &= \text{Var}(\mathbf{u}_{t|m})^{-1} \\
&= \text{Var}(\mathbf{u}_t + \mathbf{K}'_t \underline{\mathbf{L}}'_{t+1,m} \mathbf{r}_m)^{-1} \\
&= (\mathbf{M}_t - \mathbf{K}'_t \underline{\mathbf{L}}'_{t+1,m} \mathbf{N}_m \underline{\mathbf{L}}_{t+1,m} \mathbf{K}_t)^{-1}
\end{aligned} \tag{3.58}$$

because  $\text{Var}(\mathbf{u}_t) = \mathbf{M}_t$ ,  $\text{Var}(\mathbf{r}_m) = \mathbf{N}_m$  and

$$\begin{aligned}
\text{Cov}(\mathbf{u}_t, \mathbf{r}_m) &= \text{Cov}(\mathbf{F}_t^{-1} \mathbf{v}_t - \mathbf{K}'_t \mathbf{r}_t, \mathbf{r}_m) \\
&= -\mathbf{K}'_t \text{Cov}(\mathbf{r}_t, \mathbf{r}_m) \\
&= -\mathbf{K}'_t \text{Cov}(\mathbf{Z}'_{t+1} \mathbf{F}_{t+1}^{-1} \mathbf{v}_{t+1} + \mathbf{L}'_{t+1} \mathbf{r}_{t+1}, \mathbf{r}_m) \\
&= -\mathbf{K}'_t \mathbf{L}'_{t+1} \text{Cov}(\mathbf{r}_{t+1}, \mathbf{r}_m) \\
&= \vdots \\
&= -\mathbf{K}'_t \underline{\mathbf{L}}'_{t+1,m} \mathbf{N}_m
\end{aligned} \tag{3.59}$$

Thus,  $\text{MSE}(\dot{\mathbf{y}}_{t|m}) = \mathbf{M}_{t|m}$  as required.

□

### 3.3.3 State Estimate

We will now evaluate the corresponding partially interpolated state estimates  $\hat{\alpha}_{t|m}$ :

**Theorem 3.3.3.** *The interpolated MMSLE of  $\alpha_t$  conditional on the*

punctured space  $Y_m^{(t)}$  is given by

$$\dot{\alpha}_{t|m} = \hat{\alpha}_{t|m} - P_t R_{t|m} M_{t|m}^{-1} u_{t|m} \quad (3.60)$$

and its mean square error is

$$\text{MSE}_{\alpha_t}(\dot{\alpha}_{t|m}) = P_{t|m} + P_t R_{t|m} M_{t|m}^{-1} R_{t|m}' P_t \quad (3.61)$$

where

$$R_{t|m} \equiv R_t + \underline{L}_{t,m}' N_m \underline{L}_{t+1,m} K_t \quad (3.62)$$

$$\hat{\alpha}_{t|m} \equiv \hat{\alpha}_t - P_t \underline{L}_{t,m}' r_m \quad (3.63)$$

$$P_{t|m} \equiv \text{MSE}_{\alpha_t}(\hat{\alpha}_{t|m}) = P_t + P_t \underline{L}_{t,m}' N_m \underline{L}_{t,m} P_t \quad (3.64)$$

*Proof.* Similarly to the proof of Theorem 3.2.3, we have

$$\begin{aligned}
\hat{\alpha}_{t|m} &= L(\alpha_t | Y_m^{(t)}, y_t) \\
&= L(\alpha_t | Y_m^{(t)}, y_t - \dot{y}_{t|m}) \\
&= L(\alpha_t | Y_m^{(t)}) + \text{Cov}(\alpha_t, y_t - \dot{y}_{t|m}) \text{Var}(y_t - \dot{y}_{t|m})^{-1} (y_t - \dot{y}_{t|m}) \\
&= \dot{\alpha}_{t|m} + \text{Cov}(\alpha_t, y_t - \dot{y}_{t|m}) \frac{\text{MSE}(\dot{y}_{t|m})^{-1}}{\mathbf{y}_t} (y_t - \dot{y}_{t|m}) \\
&= \dot{\alpha}_{t|m} + \text{Cov}(\alpha_t, y_t - \dot{y}_{t|m}) \mathbf{u}_{t|m} \\
&= \dot{\alpha}_{t|m} + \text{Cov}(\alpha_t, M_{t|m}^{-1} \mathbf{u}_{t|m}) \mathbf{u}_{t|m} \\
&= \dot{\alpha}_{t|m} + \text{Cov}(\alpha_t, \mathbf{u}_t + \mathbf{K}_t' \underline{\mathbf{L}}_{t+1,m}' \mathbf{r}_m) M_{t|m}^{-1} \mathbf{u}_{t|m} \\
&= \dot{\alpha}_{t|m} + [\text{Cov}(\alpha_t, \mathbf{u}_t) + \text{Cov}(\alpha_t, \mathbf{r}_m) \underline{\mathbf{L}}_{t+1,m} \mathbf{K}_t'] M_{t|m}^{-1} \mathbf{u}_{t|m}
\end{aligned} \tag{3.65}$$

We have already evaluated  $\text{Cov}(\alpha_t, \mathbf{u}_t) = \mathbf{P}_t \mathbf{R}_t$ . The remaining covariance term in the expression is

$$\begin{aligned}
\text{Cov}(\alpha_t, \mathbf{r}_m) &= \text{Cov}(\alpha_t - \hat{\alpha}_t + \hat{\alpha}_t, \mathbf{r}_m) \\
&= \text{Cov}(\alpha_t - \hat{\alpha}_t, \mathbf{r}_m) + \text{Cov}(\hat{\alpha}_t, \mathbf{r}_m) \\
&= \text{Cov}(\mathbf{a}_t + \mathbf{P}_t \mathbf{r}_{t-1}, \mathbf{r}_m) \\
&= \text{Cov}(\mathbf{a}_t, \mathbf{r}_m) + \mathbf{P}_t \text{Cov}(\mathbf{r}_{t-1}, \mathbf{r}_m) \\
&= \mathbf{P}_t \underline{\mathbf{L}}_{t,m}' \mathbf{N}_m
\end{aligned} \tag{3.66}$$

using results from the proof of Theorem 3.2.3. We can thus deduce

$$\hat{\alpha}_{t|m} = \hat{\alpha}_{t|m} + P_t (R_t + \underline{L}'_{t,m} N_m \underline{L}_{t+1,m} K_t) M_{t|m}^{-1} \mathbf{u}_{t|m} \quad (3.67)$$

where the expression inside the brackets is equal to  $R_{t|m}$ . This proves the first part of Theorem 3.3.3. Proof of the second part follows directly, since the estimation error  $\hat{\alpha}_{t|m} - \alpha_t$  is independent of  $\mathbf{u}_{t|m}$ . The associated MSE matrix is

$$\begin{aligned} \text{MSE}_{\alpha_t}(\hat{\alpha}_{t|m}) &= \text{Var}(\hat{\alpha}_{t|m} - P_t R_{t|m} M_{t|m}^{-1} \mathbf{u}_{t|m} - \alpha_t) \\ &= \text{Var}(\hat{\alpha}_{t|m} - \alpha_t) + \text{Var}(P_t R_{t|m} M_{t|m}^{-1} \mathbf{u}_{t|m}) \\ &= \text{Var}(\hat{\alpha}_t - P_t \underline{L}'_{t,m} \mathbf{r}_m - \alpha_t) + P_t R_{t|m} M_{t|m}^{-1} R'_{t|m} P_t \\ &= \text{Var}(\hat{\alpha}_t - \alpha_t) + \text{Var}(P_t \underline{L}'_{t,m} \mathbf{r}_m) + P_t R_{t|m} M_{t|m}^{-1} R'_{t|m} P_t \\ &= P_t + P_t \underline{L}'_{t,m} R_m \underline{L}_{t,m} P_t + P_t R_{t|m} M_{t|m}^{-1} R'_{t|m} P_t \quad (3.68) \end{aligned}$$

which completes the proof of the theorem. □

### 3.4 Conclusions

The theorems proved in this chapter are powerful tools which further extend the usefulness of the KFS by providing us with additional methods for utilising the output quantities. We have provided a novel proof for the general case of the fundamental interpolation result, which meets the same standards of rigour as earlier proofs of the special case.

The partial-interpolation formulae can be viewed as generalisations of



the main result, and allow us greater flexibility when choosing subsets of the data on which to condition our estimates. The estimators introduced are computationally efficient to construct, as they only require the output from a single forward and backward pass of the KFS recursions.

In Chapter 4, the partial-interpolation results allow us to make full use of the KFS output from the univariate representation of a multivariate time series and construct estimators with desirable properties. These results then form the basis of the various model-based methods for estimation of missing temperature data in Chapter 6.

## Chapter 4

# The Univariate Representation of a Multivariate Time Series

### 4.1 Introduction

When a time series model is in state space form, the Kalman filter can be used to obtain filtered or smoothed estimates of the underlying states  $\alpha_t$  given the observed series. In this chapter we demonstrate how the multivariate series  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ , where  $\mathbf{y}_t = (y_{1,t}, y_{2,t}, \dots, y_{p,t})'$ , can be treated as the univariate series  $y_{1,1}, y_{2,1}, \dots, y_{p,1}, y_{1,2}, \dots, y_{p,n}$  for the purposes of filtering and smoothing. We will refer to  $\{\mathbf{y}_t\}$  as the multivariate representation (MR) and  $\{y_{i,t}\}$  as the univariate representation (UR) of the time series.

Anderson & Moore (1979) introduce the concept of decomposing the vector  $\mathbf{y}_t$  into vectors of smaller dimension, which they term sequential processing. An application of a similar method to longitudinal models is explored by Fahrmeir & Tutz (1994). Koopman & Durbin (2000) offer a more detailed treatment of the UR, but their approach only focuses on the basic state space representation, where the measurement and transition errors are uncorrelated.

We expand on their work by modifying the UR so that it can be applied

to the more general SSF of de Jong (1991). We examine the output of the UR Kalman filter and smoother in detail, and discuss how it is related to the MR filter and smoother output.

Finally, we extend the concept of the UR to any time series where several terms are observed contemporaneously. We demonstrate how this approach can be used to partition the variance of the state MLE (or, in the non-Gaussian case, the MSE of the state MMSLE) into contemporaneous and non-contemporaneous components, as a starting point for analysis of variance.

## 4.2 Motivation

Use of the UR has several advantages:

1. In the MR Kalman filter, if the single point  $y_{i,t}$  is missing, the whole vector  $\mathbf{y}_t$  needs to be treated as missing. In the UR, the contemporaneous observations  $y_{1,t}, \dots, y_{i-1,t}, y_{i+1,t}, \dots, y_{p_t,t}$  still enter the filter, thus improving the accuracy of smoothed estimates. Additionally, the ability of the UR Kalman filter to ignore individual components of the vector  $\mathbf{y}_t$  allows us to compute deletion residuals that result from removing individual points or patches of observations from the sample.
2. The initial MR state vector may contain both diffuse and non-diffuse elements. Implementation of the Kalman filter with these partially-diffuse initial conditions is more straightforward with the UR (Durbin & Koopman, 2001).
3. If the components of  $\mathbf{y}_t$  are not observed at the same time, the UR makes it possible to update the estimates as soon as new data arrives.
4. The computations required for the MR Kalman filter involve calculating  $\mathbf{F}_t^{-1}$ , the inverse of the innovation variance, for every value of  $t$ . The innovations have the same dimension as the observations, thus each step of the filter will involve inverting a  $p_t \times p_t$  matrix. In the UR, no matrix inversions are necessary, since the innovation variances  $\{F_{i,t}\}$  are scalar. Koopman & Durbin (2000) give results

purporting to show that this leads to considerable computational gains, however, despite appreciable reprogramming, it has been impossible to reproduce their results.

## 4.3 Reformulating the State Space Model

### 4.3.1 The Koopman-Durbin Approach

Koopman and Durbin (2000) specify the state space model as:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{R}_t \boldsymbol{\eta}_t \quad t = 1, \dots, n \end{aligned} \quad (4.1)$$

where  $\{\boldsymbol{\varepsilon}_t\} \sim \text{WN}(\mathbf{0}, \mathbf{H}_t)$ ,  $\{\boldsymbol{\eta}_t\} \sim \text{WN}(\mathbf{0}, \mathbf{S}_t)$  and  $\text{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\eta}_s) = \mathbf{0}$  for all  $t, s = 1, \dots, n$ . As in the general SSF, the first state  $\boldsymbol{\alpha}_1$  has (unconditional) mean  $\mathbf{a}_1$  and variance  $\mathbf{P}_1$ . If the components of the measurement error  $\boldsymbol{\varepsilon}_t$  are uncorrelated (i.e. if  $\mathbf{H}_t$  is a diagonal matrix), the model is easy to reformulate. Let  $\mathbf{Z}_{1,t}$  be the  $i^{\text{th}}$  row of the measurement matrix  $\mathbf{Z}_t$ , that is,

$$\mathbf{Z}_t = \begin{bmatrix} \mathbf{Z}_{1,t} \\ \vdots \\ \mathbf{Z}_{p_t,t} \end{bmatrix} \quad (4.2)$$

The model can be expressed in the form

$$\begin{aligned} y_{i,t} &= \mathbf{Z}_{i,t} \boldsymbol{\alpha}_{i,t} + \varepsilon_{i,t} \\ \boldsymbol{\alpha}_{i,t+1} &= \begin{cases} \mathbf{T}_t \boldsymbol{\alpha}_{p_t,t} + \mathbf{R}_t \boldsymbol{\eta}_t & \text{if } i = 1 \\ \boldsymbol{\alpha}_{i-1,t+1} & \text{if } i = 2, 3, \dots, p_{t+1} \end{cases} \end{aligned} \quad (4.3)$$

for  $t = 1, \dots, n$ , where  $\boldsymbol{\alpha}_{1,1} = \boldsymbol{\alpha}_1$ . This is the SSF of the univariate

representation of a multivariate time series.

From time  $(1, t)$  to  $(p_t, t)$ , the UR state  $\alpha_{i,t}$  stays constant; the transition matrix is the unit matrix and the transition error is zero. At each time period  $(i, t)$ , the relevant part of the state vector is measured by premultiplying  $\alpha_{i,t}$  by the rows of the measurement matrix,  $Z_{i,t}$ , and adding the measurement error  $\varepsilon_{i,t}$ . This simple sequential approach is possible because contemporaneous measurement errors are uncorrelated, thus there is no need for them to enter the model simultaneously (as the vector  $\varepsilon_t$ ).

The transition from  $\alpha_{p_t,t}$  to  $\alpha_{1,t+1}$  in the UR is identical to the transition from  $\alpha_t$  to  $\alpha_{t+1}$  in the MR, since the corresponding transition and transition error matrices are the same.

#### 4.3.2 Non-diagonal Measurement Error Variance

If the components of the measurement error are correlated, i.e.  $H_t$  is non-diagonal, the simple sequential approach is not directly applicable, because the elements of  $\varepsilon_t$  need to enter the model simultaneously. Koopman and Durbin propose augmenting the state vector  $\alpha_t$  by incorporating the error term  $\varepsilon_t$ . We adapt their method by defining

$$\begin{aligned} \tilde{Z}_t &= \begin{bmatrix} Z_t & I_{p_t} \end{bmatrix}, \quad \tilde{\alpha}_t = \begin{bmatrix} \alpha_t \\ \varepsilon_t \end{bmatrix}, \quad \tilde{T}_t = \begin{bmatrix} T_t & 0 \\ 0 & 0 \end{bmatrix} \\ \tilde{R}_t &= \begin{bmatrix} R_t & 0 \\ 0 & I_{p_t} \end{bmatrix}, \quad \tilde{\eta}_t = \begin{bmatrix} \eta_t \\ \varepsilon_{t+1} \end{bmatrix}, \quad \tilde{S}_t = \begin{bmatrix} S_t & 0 \\ 0 & H_{t+1} \end{bmatrix} \\ \tilde{a}_1 &= \begin{bmatrix} a_1 \\ 0 \end{bmatrix}, \quad \tilde{P}_1 = \begin{bmatrix} P_1 & 0 \\ 0 & H_1 \end{bmatrix} \end{aligned} \quad (4.4)$$

In this formulation, both elements required to calculate observation  $y_t$  (the state vector  $\alpha_t$  and the error  $\varepsilon_t$ ) are contained in the augmented state vector  $\tilde{\alpha}_t$ . Thus, the noise term is eliminated from the measurement equation and the MR model can be written in the form

$$\begin{aligned} \mathbf{y}_t &= \tilde{\mathbf{Z}}_t \tilde{\boldsymbol{\alpha}}_t \\ \tilde{\boldsymbol{\alpha}}_{t+1} &= \tilde{\mathbf{T}}_t \tilde{\boldsymbol{\alpha}}_t + \tilde{\mathbf{R}}_t \tilde{\boldsymbol{\eta}}_t \quad t = 1, \dots, n \end{aligned} \quad (4.5)$$

where  $\{\tilde{\boldsymbol{\eta}}_t\} \sim \text{WN}(\mathbf{0}, \tilde{\mathbf{S}}_t)$ , and the first state  $\tilde{\boldsymbol{\alpha}}_1$  has unconditional mean and variance  $\sim (\tilde{\mathbf{a}}_1, \tilde{\mathbf{P}}_1)$ , respectively.

The advantage of our parameterisation is that Koopman and Durbin define the transition error and variance of the series as

$$\tilde{\boldsymbol{\eta}}_t = \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\varepsilon}_t \end{bmatrix}, \quad \tilde{\mathbf{S}}_t = \begin{bmatrix} \mathbf{S}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_t \end{bmatrix} \quad (4.6)$$

This is incorrect; the state vector  $\tilde{\boldsymbol{\alpha}}_{t+1}$  on the LHS of the transition equation involves the term  $\boldsymbol{\varepsilon}_{t+1}$ , so this needs to appear in the associated error term on the RHS. A more serious problem is that the Koopman-Durbin formulation violates the conditions of the SSF since the state  $\tilde{\boldsymbol{\alpha}}_t$  is no longer independent of all transition errors  $\tilde{\boldsymbol{\eta}}_i$  for  $i = 1 \dots t$ , because  $\tilde{\boldsymbol{\alpha}}_t$  and  $\tilde{\boldsymbol{\eta}}_t$  both contain the component  $\boldsymbol{\varepsilon}_t$ .

Furthermore, the simple SSF (4.1) is inadequate, as the measurement and transition errors are necessarily uncorrelated. Allowing for possible correlation is essential for a wide variety of state space models (Harvey, 1989), such as the  $\max(p, q)$  representation of an  $\text{ARMA}(p, q)$  model (Pearlman, 1980; de Jong & Penzer, 2004).

### 4.3.3 A More General State Space Model

Consider the SSF as defined by de Jong and Penzer (1998):

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{G}_t \boldsymbol{\varepsilon}_t \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t \quad , \quad t = 1, \dots, n \end{aligned} \quad (4.7)$$

where  $\{\boldsymbol{\varepsilon}_t\} \sim \text{WN}(\mathbf{0}, \sigma^2 \mathbf{I}_r)$  and  $\boldsymbol{\alpha}_1$  has mean  $\mathbf{a}_1$  and variance  $\mathbf{P}_1$ . The error terms  $\mathbf{G}_t \boldsymbol{\varepsilon}_t$  and  $\mathbf{H}_t \boldsymbol{\varepsilon}_t$  have variance  $\sigma^2 \mathbf{G}_t \mathbf{G}_t'$  and  $\sigma^2 \mathbf{H}_t \mathbf{H}_t'$  respectively. We will assume  $\sigma^2 = 1$  without loss of generality, redefining  $\mathbf{G}_t$  and  $\mathbf{H}_t$  if necessary. The correlation between measurement and transition errors is made explicit by using the same  $r \times 1$  error vector  $\boldsymbol{\varepsilon}_t$  in both the measurement and the transition equation and introducing matrices  $\mathbf{G}_t, \mathbf{H}_t$ .

The correlation between contemporaneous measurement and transition errors is  $\text{Cov}(\mathbf{G}_t \boldsymbol{\varepsilon}_t, \mathbf{H}_t \boldsymbol{\varepsilon}_t) = \mathbf{G}_t \text{Var}(\boldsymbol{\varepsilon}_t) \mathbf{H}_t' = \mathbf{G}_t \mathbf{H}_t'$ . If  $\mathbf{G}_t$  and  $\mathbf{H}_t$  are orthogonal, this is equal to zero, the errors are uncorrelated and we can use the Koopman-Durbin approach to provide a UR. This becomes clearer if the term  $\mathbf{G}_t \boldsymbol{\varepsilon}_t$  is replaced by the  $p \times 1$  vector  $\boldsymbol{\zeta}_t \sim \text{WN}(\mathbf{0}, \mathbf{G}_t \mathbf{G}_t')$ , which is independent of  $\mathbf{H}_t \boldsymbol{\varepsilon}_t$ . The model is now

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\zeta}_t \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t \quad t = 1, \dots, n \end{aligned} \quad (4.8)$$

and is equivalent to the simple SSF (4.1) up to a relabelling of the system matrices.

#### 4.3.4 Correlation between Measurement Error and Transition error

If there is correlation between the measurement error and the transition error ( $\mathbf{G}_t \mathbf{H}_t' \neq \mathbf{0}$ ), the UR can be constructed by incorporating the common error term  $\boldsymbol{\varepsilon}_t$  into the state vector. The new system matrices, state vector and error term are

$$\check{\mathbf{Z}}_t = \begin{bmatrix} \mathbf{Z}_t & \mathbf{G}_t \end{bmatrix}, \quad \check{\boldsymbol{\alpha}}_t = \begin{bmatrix} \boldsymbol{\alpha}_t \\ \boldsymbol{\varepsilon}_t \end{bmatrix}, \quad \check{\mathbf{G}}_t = \mathbf{0}, \quad \check{\mathbf{T}}_t = \begin{bmatrix} \mathbf{T}_t & \mathbf{H}_t \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\check{H}_t = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_r \end{bmatrix}, \quad \check{\epsilon}_t = \epsilon_{t+1}, \quad \check{a}_1 = \begin{bmatrix} a_1 \\ \mathbf{0} \end{bmatrix}, \quad \check{P}_1 = \begin{bmatrix} P_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_r \end{bmatrix}$$

The resulting state space model is

$$\begin{aligned} y_t &= \check{Z}_t \check{\alpha}_t \\ \check{\alpha}_{t+1} &= \check{T}_t \check{\alpha}_t + \check{H}_t \check{\epsilon}_t, \quad t = 1, \dots, n \end{aligned} \quad (4.9)$$

where  $\{\check{\epsilon}_t\} \sim \text{WN}(\mathbf{0}, \mathbf{I}_q)$  and  $\check{\alpha}_1$  has mean  $\check{a}_1$  and variance  $\check{P}_1$ . Because of the change in the index of the error term,  $\check{\epsilon}_t$  is independent of  $\check{\alpha}_t$ , so the model is still in SSF. There is no measurement error, thus the UR is

$$\begin{aligned} y_{i,t} &= \check{Z}_{i,t} \check{\alpha}_{i,t} \\ \check{\alpha}_{i,t+1} &= \begin{cases} \check{T}_t \check{\alpha}_{p_t,t} + \check{H}_t \check{\epsilon}_t & \text{if } i = 1 \\ \check{\alpha}_{i-1,t+1} & \text{if } i = 2, 3, \dots, p_{t+1} \end{cases} \end{aligned} \quad (4.10)$$

where  $\check{\alpha}_{1,1} = \check{\alpha}_1$  and we define  $G_{i,t}$ ,  $\check{Z}_{i,t}$  as the  $i^{\text{th}}$  row of  $G_t$ ,  $\check{Z}_t$ , respectively:

$$\check{Z}_t = \begin{bmatrix} \check{Z}_{1,t} \\ \vdots \\ \check{Z}_{p_t,t} \end{bmatrix} = \begin{bmatrix} Z_{1,t} & G_{1,t} \\ \vdots & \vdots \\ Z_{p_t,t} & G_{p_t,t} \end{bmatrix} \quad (4.11)$$

Conceptually, this model is similar to the simple sequential UR, in that the state vector stays constant from  $(1, t)$  to  $(p_t, t)$  as the relevant parts are read off. The only difference is in the transition step, from  $(p_t, t)$  to  $(1, t+1)$ , as the error term  $\epsilon_{t+1}$  is sampled and stored in the augmented state vector  $\check{\alpha}_{i,t+1}$ .



## 4.4 Filtering

### 4.4.1 Univariate Filtering

In the previous section we establish that the measurement error can always be eliminated from the model by incorporating into the state vector. Hence, in the interest of notational simplicity, the general SSF (4.7) can be expressed as

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t \quad , \quad t = 1, \dots, n \end{aligned} \quad (4.12)$$

without loss of generality. In the absence of measurement noise, the Kalman filter equations can be simplified considerably:

$$\begin{aligned} \mathbf{v}_t &= \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t \\ \mathbf{F}_t &= \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' \\ \mathbf{K}_t &= \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' \mathbf{F}_t^{-1} \\ \mathbf{L}_t &= \mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t \\ \mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t \mathbf{v}_t \\ \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{H}_t \mathbf{H}_t' \end{aligned} \quad (4.13)$$

Note that the  $p_t \times p_t$  matrix  $\mathbf{F}_t$  needs to be inverted at each time period  $t$ . This is the most computationally expensive step in the recursions.

The univariate representation corresponding to (4.12) has state space form:

$$\begin{aligned} y_{i,t} &= \mathbf{Z}_{i,t} \boldsymbol{\alpha}_{i,t} \\ \boldsymbol{\alpha}_{i,t+1} &= \begin{cases} \mathbf{T}_t \boldsymbol{\alpha}_{p_t,t} + \mathbf{H}_t \boldsymbol{\varepsilon}_t & \text{if } i = 1 \\ \boldsymbol{\alpha}_{i-1,t+1} & \text{if } i = 2, 3, \dots, p_{t+1} \end{cases} \end{aligned} \quad (4.14)$$

where  $\mathbf{Z}_{i,t}$  is the  $i^{th}$  row of  $\mathbf{Z}_t$  and  $\boldsymbol{\alpha}_{i,t} = \boldsymbol{\alpha}_t$ . The Kalman filter recursions for the UR run forwards from  $i = 1, \dots, p_t$  for each  $t = 1, \dots, n$ :

$$\begin{aligned}
\tilde{v}_{i,t} &= y_{i,t} - \mathbf{Z}_{i,t} \tilde{\mathbf{a}}_{i,t} \\
\tilde{F}_{i,t} &= \mathbf{Z}_{i,t} \tilde{\mathbf{P}}_{i,t} \mathbf{Z}_{i,t}' \\
\tilde{\mathbf{K}}_{i,t} &= \begin{cases} \tilde{\mathbf{P}}_{i,t} \mathbf{Z}_{i,t}' \tilde{F}_{i,t}^{-1} & \text{if } i = 1, 2, \dots, p_t - 1 \\ \mathbf{T}_t \tilde{\mathbf{P}}_{p_t,t} \mathbf{Z}_{p_t,t}' \tilde{F}_{p_t,t}^{-1} & \text{if } i = p_t \end{cases} \\
\tilde{\mathbf{L}}_{i,t} &= \begin{cases} \mathbf{I} - \tilde{\mathbf{K}}_{i,t} \mathbf{Z}_{i,t} & \text{if } i = 1, 2, \dots, p_t - 1 \\ \mathbf{T}_t - \tilde{\mathbf{K}}_{p_t,t} \mathbf{Z}_{p_t,t} & \text{if } i = p_t \end{cases} \\
\tilde{\mathbf{a}}_{i,t+1} &= \begin{cases} \mathbf{T}_t \tilde{\mathbf{a}}_{p_t,t} + \tilde{\mathbf{K}}_{p_t,t} \tilde{v}_{p_t,t} & \text{if } i = 1 \\ \tilde{\mathbf{a}}_{i-1,t+1} + \tilde{\mathbf{K}}_{i-1,t+1} \tilde{v}_{i-1,t+1} & \text{if } i = 2, 3, \dots, p_{t+1} \end{cases} \\
\tilde{\mathbf{P}}_{1,t+1} &= \begin{cases} \mathbf{T}_t \tilde{\mathbf{P}}_{p_t,t} \tilde{\mathbf{L}}_{p_t,t}' + \mathbf{H}_t \mathbf{H}_t' & \text{if } i = 1 \\ \tilde{\mathbf{P}}_{i-1,t+1} \tilde{\mathbf{L}}_{i-1,t+1}' & \text{if } i = 2, 3, \dots, p_{t+1} \end{cases} \quad (4.15)
\end{aligned}$$

for  $t = 1, \dots, n$ . Koopman and Durbin identify certain cases where the innovation variance is zero in some steps of the filter recursions. This could occur, for example, if there is an observation  $y_{i,t}$  which is a linear combination of “past-contemporaneous” terms  $\{y_{1,t}, \dots, y_{i-1,t}\}$ . In that case, the Kalman gain in the UR is equal to zero ( $\tilde{\mathbf{K}}_{i,t} = 0$ ), so the state estimate and its variance do not need to be updated.

The advantage of the UR Kalman filter lies in the fact that  $\tilde{F}_{i,t}$  is a scalar, hence no matrix inversions are necessary. The output of the univariate filter is different to that of the multivariate filter; for example, if  $v_{i,t}$  is the  $i^{th}$  element of  $\mathbf{v}_t$ , we have  $\tilde{v}_{i,t} \neq v_{i,t}$ . In the following subsection we show how these terms are related.

#### 4.4.2 Filtered State Estimate

Define  $\underline{\mathbf{y}}_{i,t} = \{y_{1,t}, \dots, y_{i,t}\}$ , with  $\underline{\mathbf{y}}_{0,t} = \emptyset$  and  $\underline{\mathbf{y}}_{p_t,t} = \mathbf{y}_t$ . We will refer to  $\underline{\mathbf{y}}_{i-1,t}$  as the *contemporaneous past* of observation  $y_{i,t}$  because it consists of all the elements of  $\mathbf{y}_t$  which, in the UR, enter the model before time  $(i, t)$ . We also define  $\mathbf{Y}_{i,t} = \{\mathbf{Y}_{t-1}, \underline{\mathbf{y}}_{i,t}\}$ , with  $\mathbf{Y}_{0,t} = \mathbf{Y}_{t-1}$  and  $\mathbf{Y}_{p_t,t} = \{\mathbf{Y}_{t-1}, \mathbf{y}_t\} = \mathbf{Y}_t$ . We term  $\mathbf{Y}_{i-1,t}$  the *augmented past* of observation  $y_{i,t}$ . The UR Kalman filter produces the MMSLE of the state  $\tilde{\mathbf{a}}_{i,t}$  conditional on the augmented past  $\mathbf{Y}_{i-1,t}$ . For  $i = 1$ , this MMSLE and its associated MSE are

$$\begin{aligned}\tilde{\mathbf{a}}_{1,t} &= \mathbf{L}(\boldsymbol{\alpha}_{1,t} | \mathbf{Y}_{0,t}) = \mathbf{L}(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}) = \mathbf{a}_t \\ \tilde{\mathbf{P}}_{1,t} &= \text{MSE}(\mathbf{a}_{1,t}) = \text{MSE}(\mathbf{a}_t) = \mathbf{P}_t\end{aligned}\quad (4.16)$$

The terms  $\tilde{\mathbf{a}}_{1,t}, \tilde{\mathbf{a}}_{2,t}, \dots, \tilde{\mathbf{a}}_{p_t,t}$  are all estimates of the same state  $\boldsymbol{\alpha}_t$ . The MR state estimator  $\mathbf{a}_t$  only depends on past information ( $\mathbf{Y}_{t-1}$ ), whereas the UR estimator  $\tilde{\mathbf{a}}_{i,t}$  also takes into account the contemporaneous past ( $\underline{\mathbf{y}}_{i-1,t}$ ). As the contemporaneous observations enter the model, the variance is updated as follows:

$$\begin{aligned}\tilde{\mathbf{P}}_{i+1,t} &= \tilde{\mathbf{P}}_{i,t} \tilde{\mathbf{L}}_{i,t}' \\ &= \tilde{\mathbf{P}}_{i,t} - \tilde{\mathbf{P}}_{i,t} \mathbf{Z}_{i,t}' \tilde{\mathbf{K}}_{i,t}' \\ &= \tilde{\mathbf{P}}_{i,t} - \tilde{\mathbf{P}}_{i,t} \mathbf{Z}_{i,t}' \tilde{\mathbf{F}}_{i,t}^{-1} \tilde{\mathbf{F}}_{i,t} \tilde{\mathbf{K}}_{i,t}' \\ &= \tilde{\mathbf{P}}_{i,t} - \tilde{\mathbf{K}}_{i,t} \text{Var}(\tilde{v}_{i,t}) \tilde{\mathbf{K}}_{i,t}' \\ &= \tilde{\mathbf{P}}_{i,t} - \text{Var}(\tilde{\mathbf{K}}_{i,t} \tilde{v}_{i,t})\end{aligned}\quad (4.17)$$

Thus, the MSE matrix  $\tilde{\mathbf{P}}_{i+1,t}$  is smaller than  $\tilde{\mathbf{P}}_{i,t}$ , in the sense that the difference  $\tilde{\mathbf{P}}_{i,t} - \tilde{\mathbf{P}}_{i+1,t}$  is equal to a variance (i.e. positive semi-definite) matrix. Hence, each diagonal element of  $\tilde{\mathbf{P}}_{i+1,t}$  is smaller than or equal to the corresponding element of  $\tilde{\mathbf{P}}_{i,t}$ . They are equal only in the situation where  $y_{i,t}$  is a linear function of past observations  $\mathbf{Y}_{i-1,t}$ , in which case the Kalman gain is zero and  $\tilde{\mathbf{P}}_{i+1,t} = \tilde{\mathbf{P}}_{i,t}$ . In general,

the strict inequality holds, which means that the MSE of  $\tilde{\mathbf{a}}_{i+1,t}$  as an estimator of the state  $\boldsymbol{\alpha}_t$  is smaller than the MSE of  $\tilde{\mathbf{a}}_{i,t}$  as an estimator of the same state. This is intuitively reasonable; if the state is constant and there is no measurement noise, the estimate of the state is improved as more data enters the KF.

We can also write

$$\begin{aligned}
\tilde{\mathbf{P}}_{i+1,t} &= \tilde{\mathbf{P}}_{i,t} \tilde{\mathbf{L}}_{i,t}' \\
&= \tilde{\mathbf{P}}_{i-1,t} \tilde{\mathbf{L}}_{i-1,t}' \tilde{\mathbf{L}}_{i,t}' \\
&= \vdots \\
&= \tilde{\mathbf{P}}_{1,t} \tilde{\mathbf{L}}_{1,t}' \\
&= \mathbf{P}_t \tilde{\mathbf{L}}_{i,t}'
\end{aligned} \tag{4.18}$$

where  $\tilde{\mathbf{L}}_{i,t}' = \tilde{\mathbf{L}}_{1,t}' \dots \tilde{\mathbf{L}}_{i,t}'$ . Repeated application of identity (4.18) produces  $\tilde{\mathbf{P}}_{p_t,t} = \mathbf{P}_t \tilde{\mathbf{L}}_{p_t,t}'$ , which can be substituted into (4.15) to yield

$$\mathbf{P}_{t+1} = \tilde{\mathbf{P}}_{1,t+1} = \mathbf{T}_t \mathbf{P}_t \tilde{\mathbf{L}}_{p_t,t}' + \mathbf{H}_t \mathbf{H}_t' \tag{4.19}$$

Expression (4.19) is equivalent to the MR recursion

$$\mathbf{P}_{t+1} = \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{H}_t \mathbf{H}_t' \tag{4.20}$$

This appears to suggest that the product of the accumulated  $\tilde{\mathbf{L}}_{i,t}$  matrices from  $(1, t)$  to  $(p_t, t)$  in the UR filter is equal to the corresponding matrix in the MR filter, that is,

$$\tilde{\mathbf{L}}_{p_t,t} = \mathbf{L}_t \tag{4.21}$$

However, (4.21) cannot be deduced directly from (4.19) and (4.20) be-

cause the matrix  $\mathbf{T}_t$  is not necessarily invertible. In the following chapter we prove a general result of which (4.21) is a special case.

#### 4.4.3 Innovations

The innovations  $\mathbf{v}_t$  from the MR can be reconstructed from the UR output, since

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t = \mathbf{y}_t - \mathbf{Z}_t \tilde{\mathbf{a}}_{1,t} \quad (4.22)$$

Using the notation  $\mathbf{v}_t = (v_{1,t}, \dots, v_{p_t,t})'$ , we can see that  $\tilde{v}_{1,t} = v_{1,t}$ . For  $i > 1$ , we have

$$\tilde{v}_{i,t} = y_{i,t} - L(y_{i,t} | \mathbf{Y}_{i-1,t}) \quad (4.23)$$

The  $i^{th}$  observation at time  $t$  can be written in terms of the corresponding UR innovation as  $y_{i,t} = \tilde{v}_{i,t} + L(y_{i,t} | \mathbf{Y}_{i-1,t})$ . The term  $L(y_{i,t} | \mathbf{Y}_{i-1,t})$  is a function of  $\mathbf{Y}_{i-1,t}$  and the system matrices only. Thus, when  $(\mathbf{Y}_{i-1,t}, \tilde{v}_{i,t})$  is fixed,  $\mathbf{Y}_{i,t}$  is fixed and vice versa. The same is true for  $(\mathbf{Y}_{i-2,t}, \tilde{v}_{i-1,t}, \tilde{v}_{i,t})$  and  $\mathbf{Y}_{i,t}$ . Applying this argument recursively, we can see that if  $(\mathbf{Y}_{t-1}, \tilde{v}_{1,t}, \dots, \tilde{v}_{i,t})$  is fixed, then  $\mathbf{Y}_{i,t}$  is fixed and vice versa. Thus, equation (4.23) can be written as

$$\tilde{v}_{i,t} = y_{i,t} - L(y_{i,t} | \mathbf{Y}_{t-1}, \tilde{v}_{1,t}, \dots, \tilde{v}_{i-1,t}) \quad (4.24)$$

In direct analogy to the MR Kalman filter, the innovations from the UR  $\tilde{v}_{1,t}, \dots, \tilde{v}_{i-1,t}$  have zero mean, are pairwise independent and also independent of  $\mathbf{Y}_{t-1}$ , thus

$$\begin{aligned}
\tilde{v}_{i,t} &= y_{i,t} - L(y_{i,t} | \mathbf{Y}_{t-1}, \tilde{v}_{1,t}, \dots, \tilde{v}_{i-1,t}) \\
&= y_{i,t} - L(y_{i,t} | \mathbf{Y}_{t-1}) - \sum_{k=1}^{i-1} \text{Cov}(y_{i,t}, \tilde{v}_{k,t}) \text{Var}(\tilde{v}_{k,t})^{-1} \{\tilde{v}_{k,t} - E(\tilde{v}_{k,t})\} \\
&= v_{i,t} - \sum_{k=1}^{i-1} \text{Cov}(y_{i,t}, \tilde{v}_{k,t}) \tilde{F}_{k,t}^{-1} \tilde{v}_{k,t}
\end{aligned} \tag{4.25}$$

Since  $\alpha_{i,t} = \alpha_{k,t}$ , the covariance term on the RHS of (4.25) is

$$\begin{aligned}
\text{Cov}(y_{i,t}, \tilde{v}_{k,t}) &= \text{Cov}[\mathbf{Z}_{i,t} \alpha_{i,t}, \mathbf{Z}_{k,t} (\alpha_{k,t} - \tilde{\mathbf{a}}_{k,t})] \\
&= \mathbf{Z}_{i,t} \text{Cov}(\alpha_{k,t}, \alpha_{k,t} - \tilde{\mathbf{a}}_{k,t}) \mathbf{Z}_{k,t}'
\end{aligned} \tag{4.26}$$

where

$$\begin{aligned}
&\text{Cov}(\alpha_{k,t}, \alpha_{k,t} - \tilde{\mathbf{a}}_{k,t}) \\
&= E[\alpha_{k,t} (\alpha_{k,t} - \tilde{\mathbf{a}}_{k,t})'] \\
&= E[(\alpha_{k,t} - \tilde{\mathbf{a}}_{k,t})(\alpha_{k,t} - \tilde{\mathbf{a}}_{k,t})'] + E[\tilde{\mathbf{a}}_{k,t} (\alpha_{k,t} - \tilde{\mathbf{a}}_{k,t})'] \\
&= \text{MSE}_{\alpha_{k,t}}(\tilde{\mathbf{a}}_{k,t}) + E[L\{\tilde{\mathbf{a}}_{k,t} (\alpha_{k,t} - \tilde{\mathbf{a}}_{k,t})' | \mathbf{Y}_{k-1,t}\}] \\
&= \tilde{\mathbf{P}}_{k,t} + E[\tilde{\mathbf{a}}_{k,t} \{L(\alpha_{k,t} | \mathbf{Y}_{k-1,t}) - \tilde{\mathbf{a}}_{k,t}\}'] \\
&= \tilde{\mathbf{P}}_{k,t} + E[\tilde{\mathbf{a}}_{k,t} \{\tilde{\mathbf{a}}_{k,t} - \tilde{\mathbf{a}}_{k,t}\}'] \\
&= \tilde{\mathbf{P}}_{k,t}
\end{aligned} \tag{4.27}$$

Substituting into (4.25), we obtain

$$\begin{aligned}
\tilde{v}_{i,t} &= v_{i,t} - \sum_{k=1}^{i-1} \mathbf{Z}_{i,t} \tilde{\mathbf{P}}_{k,t} \mathbf{Z}_{k,t}' \tilde{F}_{k,t}^{-1} \tilde{v}_{k,t} \\
\Leftrightarrow v_{i,t} &= \tilde{v}_{i,t} + \sum_{k=1}^{i-1} \mathbf{Z}_{i,t} \tilde{\mathbf{K}}_{k,t} \tilde{v}_{k,t}
\end{aligned} \tag{4.28}$$

Thus, the MR innovation  $v_{i,t}$  is equal to  $\tilde{v}_{i,t}$  plus a weighted sum of past-contemporaneous UR innovations.

Define the vector of UR innovations  $\tilde{\mathbf{v}}_t = (\tilde{v}_{1,t}, \dots, \tilde{v}_{p_t,t})'$ , which has the same dimension as the MR innovation vector  $\mathbf{v}_t$ . The relation between  $\tilde{\mathbf{v}}_t$  and  $\mathbf{v}_t$  can be expressed as

$$\mathbf{v}_t = \tilde{\mathbf{Q}}_t \tilde{\mathbf{v}}_t \quad (4.29)$$

where

$$\tilde{\mathbf{Q}}_t = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \mathbf{Z}_{2,t} \tilde{\mathbf{K}}_{1,t} & 1 & 0 & \cdots & 0 \\ \mathbf{Z}_{3,t} \tilde{\mathbf{K}}_{1,t} & \mathbf{Z}_{3,t} \tilde{\mathbf{K}}_{2,t} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_{p_t,t} \tilde{\mathbf{K}}_{1,t} & \mathbf{Z}_{p_t,t} \tilde{\mathbf{K}}_{2,t} & \mathbf{Z}_{p_t,t} \tilde{\mathbf{K}}_{3,t} & \cdots & 1 \end{pmatrix} \quad (4.30)$$

This is analogous to the expression  $\mathbf{y} = \mathbf{Q}\mathbf{v}$  from the technical introduction, as will be discussed in the following subsection.

#### 4.4.4 Innovation Variance

Let  $\{\mathbf{F}_t\}_{ij}$  be the  $(i, j)^{th}$  element of the innovation variance matrix  $\mathbf{F}_t$ . Its diagonal terms can be computed by taking the variance on both sides of (4.28), because the terms on the right hand side are uncorrelated. Let  $F_{i,t} = \{\mathbf{F}_t\}_{ii}$  denote the diagonal terms. We have

$$\begin{aligned} \text{Var}(v_{i,t}) &= \text{Var}(\tilde{v}_{i,t}) + \sum_{k=1}^{i-1} \text{Var}(\mathbf{Z}_{i,t} \tilde{\mathbf{K}}_{k,t} \tilde{v}_{k,t}) \\ \Rightarrow F_{i,t} &= \tilde{F}_{i,t} + \sum_{k=1}^{i-1} \mathbf{Z}_{i,t} \tilde{\mathbf{K}}_{k,t} \tilde{F}_{k,t} \tilde{\mathbf{K}}_{k,t} \mathbf{Z}_{i,t}' \end{aligned} \quad (4.31)$$

where  $\tilde{F}_{k,t}$  and  $\tilde{\mathbf{K}}_{k,t}$  are from the univariate filter. All the terms inside the sum in (4.31) are non-negative. This implies

$$F_{i,t} \geq \tilde{F}_{i,t} \quad \forall i = 1, \dots, p_t \quad (4.32)$$

As expected, the innovation variance decreases as more past contemporaneous data enters the KF.

The non-diagonal terms can also be obtained using (4.28). Consider the case where  $i > j$ :

$$\begin{aligned} \{\mathbf{F}_t\}_{ij} &= \text{Cov}(v_{i,t}, v_{j,t}) \\ &= \text{E}(v_{i,t}v'_{j,t}) - \text{E}(v_{i,t})\text{E}(v_{j,t})' \\ &= \text{E} \left[ \left( \tilde{v}_{i,t} + \sum_{k=1}^{i-1} \mathbf{Z}_{i,t} \tilde{\mathbf{K}}_{k,t} \tilde{v}_{k,t} \right) \left( \tilde{v}_{j,t} + \sum_{l=1}^{j-1} \mathbf{Z}_{j,t} \tilde{\mathbf{K}}_{l,t} \tilde{v}_{l,t} \right)' \right] \end{aligned} \quad (4.33)$$

By the properties of innovations, only terms of the form  $\tilde{v}_{k,t}^2$  will have non-zero expectation, thus

$$\begin{aligned} \{\mathbf{F}_t\}_{ij} &= \text{E} \left[ \mathbf{Z}_{i,t} \tilde{\mathbf{K}}_{j,t} \tilde{v}_{j,t}^2 \right] + \text{E} \left[ \sum_{k=1}^{j-1} \mathbf{Z}_{i,t} \tilde{\mathbf{K}}_{k,t} \tilde{v}_{k,t}^2 \tilde{\mathbf{K}}_{k,t}' \mathbf{Z}_{j,t}' \right] \\ &= \mathbf{Z}_{i,t} \tilde{\mathbf{K}}_{j,t} \tilde{F}_{j,t} + \sum_{k=1}^{j-1} \mathbf{Z}_{i,t} \tilde{\mathbf{K}}_{k,t} \tilde{F}_{k,t} \tilde{\mathbf{K}}_{k,t}' \mathbf{Z}_{j,t}' \\ &= \mathbf{Z}_{i,t} \tilde{\mathbf{P}}_{j,t} \mathbf{Z}_{j,t} + \sum_{k=1}^{j-1} \mathbf{Z}_{i,t} \tilde{\mathbf{K}}_{k,t} \tilde{F}_{k,t} \tilde{\mathbf{K}}_{k,t}' \mathbf{Z}_{j,t}' \end{aligned} \quad (4.34)$$

When  $i = j$ , the term  $\mathbf{Z}_{i,t} \tilde{\mathbf{P}}_{j,t} \mathbf{Z}_{j,t}$  is equal to  $\tilde{F}_{i,t}$ , so (4.34) is just a special case of (4.31). Finally, when  $i < j$  we can use  $\{\mathbf{F}_t\}_{ij} = \{\mathbf{F}_t\}_{ji}$ .

Expression (4.34) can be written more concisely by exploiting the rela-



tion between  $\mathbf{v}_t$  and  $\tilde{\mathbf{v}}_t$ . Define the matrix  $\tilde{\mathbf{F}}_t = \text{Var}(\tilde{\mathbf{v}}_t)$ . The components of  $\tilde{\mathbf{v}}_t$  are uncorrelated, thus  $\tilde{\mathbf{F}}_t = \text{diag}\{\tilde{F}_{1,t}, \dots, \tilde{F}_{p_t,t}\}$ . Equation (4.29) implies

$$\begin{aligned}
\mathbf{F}_t &= \text{Var}(\mathbf{v}_t) \\
&= \text{Var}(\tilde{\mathbf{Q}}_t \tilde{\mathbf{v}}_t) \\
&= \tilde{\mathbf{Q}}_t \text{Var}(\tilde{\mathbf{v}}_t) \tilde{\mathbf{Q}}_t' \\
&= \tilde{\mathbf{Q}}_t \tilde{\mathbf{F}}_t \tilde{\mathbf{Q}}_t'
\end{aligned} \tag{4.35}$$

This is analogous to the expression  $\mathbf{\Sigma} = \mathbf{Q}\mathbf{F}\mathbf{Q}'$  from the technical introduction. The MR Kalman filter takes  $n$  steps to implicitly compute the Cholesky decomposition of the  $n \times n$ -block variance matrix  $\mathbf{\Sigma}$ , thus avoiding inverting  $\mathbf{\Sigma}$  directly. Similarly, the UR KF takes  $p_t$  steps to implicitly compute the Cholesky decomposition of the  $p_t \times p_t$  variance matrix  $\mathbf{F}_t$ , thus avoiding inverting  $\mathbf{F}_t$  directly.

## 4.5 Smoothing

### 4.5.1 Univariate Smoothing

The KFS recursions for the UR (4.14) run backwards from  $i = p_t, \dots, 1$  for each  $t = n, \dots, 1$

$$\begin{aligned}
\tilde{u}_{i,t} &= \tilde{F}_{i,t}^{-1} \tilde{v}_{i,t} - \tilde{\mathbf{K}}_{i,t}' \tilde{\mathbf{r}}_{i,t} \\
\tilde{M}_{i,t} &= \tilde{F}_{i,t}^{-1} + \tilde{\mathbf{K}}_{i,t}' \tilde{\mathbf{N}}_{i,t} \tilde{\mathbf{K}}_{i,t} \\
\tilde{\mathbf{r}}_{i-1,t} &= \begin{cases} \mathbf{Z}_{i,t}' \tilde{u}_{i,t} + \tilde{\mathbf{r}}_{i,t} & \text{if } i = 1, 2, \dots, p_t - 1 \\ \mathbf{Z}_{p_t,t}' \tilde{u}_{p_t,t} + \mathbf{T}_t' \tilde{\mathbf{r}}_{p_t,t} & \text{if } i = p_t \end{cases} \\
\tilde{\mathbf{N}}_{i-1,t} &= \mathbf{Z}_{i,t}' \tilde{F}_{i,t}^{-1} \mathbf{Z}_{i,t} + \tilde{\mathbf{L}}_{i,t}' \tilde{\mathbf{N}}_{i,t} \tilde{\mathbf{L}}_{i,t}
\end{aligned} \tag{4.36}$$

where we define  $\tilde{\mathbf{r}}_{0,t} \equiv \tilde{\mathbf{r}}_{p_{t-1},t-1}$  and  $\tilde{\mathbf{N}}_{0,t} \equiv \tilde{\mathbf{N}}_{p_{t-1},t-1}$  for notational convenience. The recursions are initialised with  $\tilde{\mathbf{r}}_{p_n,n} = 0$  and  $\tilde{\mathbf{N}}_{p_n,n} = 0$ . Alternatively, we can use the form

$$\tilde{\mathbf{r}}_{i-1,t} = \mathbf{Z}'_{i,t} \tilde{\mathbf{F}}_{i,t}^{-1} \tilde{\mathbf{v}}_{i,t} + \tilde{\mathbf{L}}'_{i,t} \tilde{\mathbf{r}}_{i,t} \quad (4.37)$$

The smoothed state estimates and their associated MSE are

$$\hat{\boldsymbol{\alpha}}_{i,t} \equiv \mathbf{L}(\boldsymbol{\alpha}_{i,t} | \mathbf{Y}_n) = \tilde{\mathbf{a}}_{i,t} + \tilde{\mathbf{P}}_{i,t} \tilde{\mathbf{r}}_{i-1,t} \quad (4.38)$$

$$\text{MSE}_{\boldsymbol{\alpha}_{i,t}}(\hat{\boldsymbol{\alpha}}_{i,t}) = \tilde{\mathbf{P}}_{i,t} - \tilde{\mathbf{P}}_{i,t} \tilde{\mathbf{N}}_{i-1,t} \tilde{\mathbf{P}}_{i,t} \quad (4.39)$$

All the UR state estimates at time  $t$  are equal, since neither the data taken into account nor the state itself changes:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_{p_t,t} &= \tilde{\mathbf{a}}_{p_t,t} + \tilde{\mathbf{P}}_{p_t,t} \tilde{\mathbf{r}}_{p_t-1,t} \\ &= \tilde{\mathbf{a}}_{p_t-1,t} + \tilde{\mathbf{K}}_{p_t-1,t} \tilde{\mathbf{v}}_{p_t-1,t} + \tilde{\mathbf{P}}_{p_t-1,t} \tilde{\mathbf{L}}'_{p_t-1,t} \tilde{\mathbf{r}}_{p_t-1,t} \\ &= \tilde{\mathbf{a}}_{p_t-1,t} + \tilde{\mathbf{P}}_{p_t-1,t} \tilde{\mathbf{Z}}'_{p_t-1,t} \tilde{\mathbf{F}}_{p_t-1,t}^{-1} \tilde{\mathbf{v}}_{p_t-1,t} + \tilde{\mathbf{P}}_{p_t-1,t} \tilde{\mathbf{L}}'_{p_t-1,t} \tilde{\mathbf{r}}_{p_t-1,t} \\ &= \tilde{\mathbf{a}}_{p_t-1,t} + \tilde{\mathbf{P}}_{p_t-1,t} \tilde{\mathbf{r}}_{p_t-2,t} \\ &= \hat{\boldsymbol{\alpha}}_{p_t-1,t} \\ &= \vdots \\ &= \hat{\boldsymbol{\alpha}}_{1,t} \end{aligned} \quad (4.40)$$

Trivially, the smoothed state estimates from the MR and UR KFS are the same:  $\hat{\boldsymbol{\alpha}}_{i,t} = \mathbf{L}(\boldsymbol{\alpha}_{i,t} | \mathbf{Y}_n) = \mathbf{L}(\boldsymbol{\alpha}_t | \mathbf{Y}_n) = \hat{\boldsymbol{\alpha}}_t$ , for  $i = 1, \dots, p_t$ . We can deduce

$$\begin{aligned}
& \hat{\tilde{\alpha}}_{1,t} = \tilde{\alpha}_t \\
\Leftrightarrow & \tilde{a}_{1,t} + \tilde{P}_{1,t} \tilde{r}_{p_{t-1},t-1} = a_t + P_t r_{t-1} \\
\Leftrightarrow & a_t + P_t \tilde{r}_{p_{t-1},t-1} = a_t + P_t r_{t-1} \\
\Leftrightarrow & \tilde{r}_{p_{t-1},t-1} = r_{t-1}
\end{aligned} \tag{4.41}$$

because the (positive semi-definite) matrix  $P_{t+1}$  is invertible. This result implies that  $\tilde{r}_{p_t,t}$  and  $r_t$  are equal for all values of  $t$ , as are their variance matrices  $\tilde{N}_{p_t,t}$  and  $N_t$ . This is intuitively reasonable:  $\tilde{r}_{p_t,t}$  and  $\tilde{N}_{p_t,t}$  are dependent only on the future of the series, so they will be equal to the equivalent MR quantities when we do not consider contemporaneous data. This occurs at the “end” of each vector  $y_t$ , in the same way that  $\tilde{a}_{i,t}$  and  $\tilde{P}_{i,t}$  only depend on the past, so they will be equal to the corresponding MR quantities when  $i = 1$ , at the “beginning” of  $y_t$ . This fact allows us to express the MR smoothations and their variance in terms of the output from the UR KFS:

$$\begin{aligned}
u_t &= F_t^{-1} v_t - K_t' r_t \\
&= (\tilde{Q}_t \tilde{F}_t \tilde{Q}_t')^{-1} \tilde{Q}_t \tilde{v}_t - K_t' \tilde{r}_{p_t,t} \\
&= \tilde{Q}_t'^{-1} \tilde{F}_t^{-1} \tilde{Q}_t^{-1} \tilde{Q}_t \tilde{v}_t - K_t' \tilde{r}_{p_t,t} \\
&= \tilde{Q}_t'^{-1} \tilde{F}_t^{-1} \tilde{v}_t - K_t' \tilde{r}_{p_t,t}
\end{aligned} \tag{4.42}$$

$$\begin{aligned}
M_t &= F_t^{-1} + K_t' N_t K_t \\
&= \tilde{Q}_t'^{-1} \tilde{F}_t^{-1} \tilde{Q}_t^{-1} + K_t' \tilde{N}_{p_t,t} K_t
\end{aligned} \tag{4.43}$$

where  $K_t$  can be written as

$$\begin{aligned}
K_t &= T_t P_t Z_t' F_t^{-1} \\
&= T_t \tilde{P}_{1,t} Z_t' (\tilde{Q}_t \tilde{F}_t \tilde{Q}_t')^{-1} \\
&= T_t \tilde{P}_{1,t} Z_t' \tilde{Q}_t'^{-1} \tilde{F}_t^{-1} \tilde{Q}_t^{-1}
\end{aligned} \tag{4.44}$$

### 4.5.2 Deletion Residuals

In many practical applications, such as detecting outliers or structural breaks in a time series, it is useful to treat  $y_{i,t}$  as a missing observation and estimate it from the rest of the data (de Jong & Penzer, 1998; Proietti, 2003; Penzer, 2007). We are particularly interested in the residual obtained by taking the difference between this estimate and the true value of  $y_{i,t}$ . This approach can also be used to assess the quality of a given dataset; if the observation  $y_{i,t}$  is unusual, its deletion residual tends to be large.

Let  $\mathbf{y}_{\setminus i,t}$  be  $\mathbf{Y}_n$  excluding  $y_{i,t}$  and also let  $\mathbf{y}_{t\setminus i}$  be  $\mathbf{y}_t$  excluding  $y_{i,t}$ . The UR KFS allows us to construct several different deletion residuals, depending on which part of the data we wish to condition on.

#### Past-only

If we only consider data up to time  $t - 1$ , the deletion residual is in fact the innovation  $v_{i,t}$  obtained through the MR KF:

$$y_{i,t} - L(y_{i,t}|\mathbf{Y}_{t-1}) = v_{i,t} = y_{i,t} - \mathbf{Z}_{i,t}\mathbf{a}_t \quad (4.45)$$

By definition, its variance is equal to  $F_{i,t}$ .

#### Present-only

If we condition on the data  $\mathbf{y}_{t\setminus i}$ , the resulting residual is

$$y_{i,t} - L(y_{i,t}|\mathbf{y}_{t\setminus i}) \quad (4.46)$$

This can be computed by putting the series  $y_{1,t}, \dots, y_{p,t}$  through the UR KFS. The forward recursion must be initialised with the unconditional mean of  $\mathbf{a}_t$  and its variance. Let these be  $\mathbf{a}_t^+$  and  $\mathbf{P}_t^+$  respectively. They

can be obtained through the recursions

$$\begin{aligned} \mathbf{a}_{t+1}^+ &= \mathbb{E}(\mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t) = \mathbf{T}_t \mathbf{a}_t^+ \\ \mathbf{P}_{t+1}^+ &= \text{Var}(\mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t) = \mathbf{T}_t \mathbf{P}_t^+ \mathbf{T}_t' + \mathbf{H}_t \mathbf{H}_t' \end{aligned} \quad (4.47)$$

for  $t = 1, \dots, n$ , initialised with  $\mathbf{a}_1^+ = \mathbf{a}_1$ ,  $\mathbf{P}_1^+ = \mathbf{P}_1$ . Once the forward pass of the KF is complete, the UR KFS will run backwards from  $y_{p_t, t}$ . The backward recursion is initialised with  $\tilde{\mathbf{r}}_{p_t, t}^+ = 0$  and  $\tilde{\mathbf{N}}_{p_t, t}^+ = 0$ . We thus obtain the smoothation  $\tilde{u}_{i, t}^+$  and its variance  $\tilde{M}_{i, t}^+$ . The resulting deletion residual is

$$y_{i, t} - \mathbb{L}(y_{i, t} | \mathbf{y}_{t \setminus i}) = (\tilde{M}_{i, t}^+)^{-1} \tilde{u}_{i, t}^+ \quad (4.48)$$

and has variance equal to  $(\tilde{M}_{i, t}^+)^{-1}$ .

### Past and present

This is the estimate we obtain if we take into account all the data up to time  $t$ , excluding  $y_{i, t}$ . To compute it, we start by running the UR Kalman filter recursions on the series  $y_{1, 1}, \dots, y_{p_t, t}$ . Once this is complete, we smooth backwards from  $y_{p_t, t}$ , initialising the KFS with  $\tilde{\mathbf{r}}_{p_t, t}^- = \mathbf{0}$  and  $\tilde{\mathbf{N}}_{p_t, t}^- = \mathbf{0}$ . This will yield the smoothation  $\tilde{u}_{i, t}^-$  and its variance  $\tilde{M}_{i, t}^-$ , which in turn allows us to construct the deletion residual

$$y_{i, t} - \mathbb{L}(y_{i, t} | \mathbf{Y}_{t-1}, \mathbf{y}_{t \setminus i}) = (\tilde{M}_{i, t}^-)^{-1} \tilde{u}_{i, t}^- \quad (4.49)$$

which has variance equal to  $(\tilde{M}_{i, t}^-)^{-1}$ .

A better method is to use the partial-interpolation formulae from Chapter 3 to construct the “past-and-present” residuals for the whole sample in a single pass of the KFS. Applying Theorem 3.3.2, we can write

$$y_{i,t} - \mathbb{L}(y_{i,t} | \mathbf{Y}_{t-1}, \mathbf{y}_{t \setminus i}) = \tilde{M}_{i,t|p_t,t}^{-1} \tilde{u}_{i,t|p_t,t} \quad (4.50)$$

where

$$\begin{aligned} \tilde{u}_{i,t|p_t,t} &\equiv \tilde{u}_{i,t} + \tilde{\mathbf{K}}'_{i,t} \tilde{\mathbf{L}}'_{i+1,t} \dots \tilde{\mathbf{L}}'_{p_t,t} \tilde{\mathbf{r}}_{p_t,t} \\ \tilde{M}_{i,t|p_t,t} &\equiv \tilde{M}_{i,t} - \tilde{\mathbf{K}}'_{i,t} \tilde{\mathbf{L}}'_{i+1,t} \dots \tilde{\mathbf{L}}'_{p_t,t} \tilde{\mathbf{N}}_{p_t,t} \tilde{\mathbf{L}}_{p_t,t} \dots \tilde{\mathbf{L}}_{i+1,t} \tilde{\mathbf{K}}_{i,t} \end{aligned} \quad (4.51)$$

The quantities  $\tilde{u}_{i,t|p_t,t}, \tilde{M}_{i,t|p_t,t}$  are the UR equivalent of  $\mathbf{u}_{t|m}, \mathbf{M}_{t|m}$  in the notation of Chapter 3.

### Past and future

If we condition on the entire sample apart from data at time  $t$ , we obtain the interpolated estimate of  $y_{i,t}$  from the MR KFS. The resulting deletion residual is

$$y_{i,t} - \mathbb{L}(y_{i,t} | \mathbf{Y}_n^{(t)}) = \{\mathbf{M}_t^{-1} \mathbf{u}_t\}_i \quad (4.52)$$

that is, the  $i^{th}$  element of the vector  $\mathbf{M}_t^{-1} \mathbf{u}_t$ . Its variance is  $M_{i,t}^{-1}$ , the  $i^{th}$  diagonal element of  $\mathbf{M}_t^{-1}$ .

### Past, present and future

Finally, we can condition on the entire sample excluding the point  $y_{i,t}$ . This estimate can be obtained from the UR KFS, with the recursions described in (4.36). This yields the residual

$$y_{i,t} - \mathbb{L}(y_{i,t} | \mathbf{y}_{\setminus i,t}) = \tilde{M}_{i,t}^{-1} \tilde{u}_{i,t} \quad (4.53)$$

with variance  $\tilde{M}_{i,t}^{-1}$ .

## 4.6 Contemporaneous Observations

### 4.6.1 Introduction

In the previous section we consider a univariate representation of a multivariate time series. This representation has the property that some of the elements of the univariate series are observed simultaneously, even though they are brought in one at a time. Specifically, at each time period  $t$ , the consecutive observations  $y_{1,t}, \dots, y_{p_t,t}$  are observed simultaneously. The UR Kalman filter recursions illustrate how, if the point of interest is  $y_{i,t}$ , the estimate of the state  $\alpha_t$  can be improved by conditioning on the data contemporaneous to  $y_{i,t}$ . This approach can be generalised to all time series with the same property.

### 4.6.2 The General Model

Consider the time series  $\{\mathbf{y}_t\}$ ,  $t = 1, 2, \dots$  which has the general SSF

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \alpha_t + \mathbf{G}_t \varepsilon_t \\ \alpha_{t+1} &= \mathbf{T}_t \alpha_t + \mathbf{H}_t \varepsilon_t \quad , \quad t = 1, \dots, n \end{aligned} \quad (4.54)$$

Now assume that data  $\mathbf{y}_{s-r}, \dots, \mathbf{y}_s$  are observed simultaneously. These represent the “present” of the series, whereas  $\mathbf{Y}_{s-r-1}$  represents the “past”. We are interested in how the past and present of the series contribute to the forecast of the current state  $\alpha_s$  and its MSE.

We must first begin by considering the covariance between the current state  $\alpha_s$  and past innovations  $\mathbf{v}_{s-j}$ :

**Lemma 4.6.1.**

$$\text{Cov}(\alpha_s, \mathbf{v}_{s-j}) = \underline{\mathbf{T}}_{s-j+1, s-1} \mathbf{K}_{s-j} \mathbf{F}_{s-j} \quad (4.55)$$

for  $j = 1, 2, \dots$  and

*Proof.* By induction on  $j$ . For  $j = 1$  we have

$$\begin{aligned}
& \text{Cov}(\boldsymbol{\alpha}_s, \mathbf{v}_{s-1}) \\
&= \text{Cov}(\mathbf{T}_{s-1}\boldsymbol{\alpha}_{s-1} + \mathbf{H}_{s-1}\boldsymbol{\varepsilon}_{s-1}, \mathbf{y}_{s-1} - \mathbf{Z}_{s-1}\mathbf{a}_{s-1}) \\
&= \text{Cov}(\mathbf{T}_{s-1}\boldsymbol{\alpha}_{s-1} + \mathbf{H}_{s-1}\boldsymbol{\varepsilon}_{s-1}, \mathbf{Z}_{s-1}\boldsymbol{\alpha}_{s-1} + \mathbf{G}_{s-1}\boldsymbol{\varepsilon}_{s-1} - \mathbf{Z}_{s-1}\mathbf{a}_{s-1}) \\
&= \text{Cov}\{\mathbf{T}_{s-1}\boldsymbol{\alpha}_{s-1}, \mathbf{Z}_{s-1}(\boldsymbol{\alpha}_{s-1} - \mathbf{a}_{s-1})\} + \text{Cov}(\mathbf{H}_{s-1}\boldsymbol{\varepsilon}_{s-1}, \mathbf{G}_{s-1}\boldsymbol{\varepsilon}_{s-1}) \\
&= \mathbf{T}_{s-1}\text{Cov}(\boldsymbol{\alpha}_{s-1}, \boldsymbol{\alpha}_{s-1} - \mathbf{a}_{s-1})\mathbf{Z}_{s-1}' + \mathbf{H}_{s-1}\mathbf{G}_{s-1}' \\
&= \mathbf{T}_{s-1}\mathbf{P}_{s-1}\mathbf{Z}_{s-1}' + \mathbf{H}_{s-1}\mathbf{G}_{s-1}' \\
&= \mathbf{K}_{s-1}\mathbf{F}_{s-1}
\end{aligned} \tag{4.56}$$

For  $j > 1$ , assuming the statement is true for all integers up to  $j - 1$ , the inductive step is

$$\begin{aligned}
\text{Cov}(\boldsymbol{\alpha}_t, \mathbf{v}_{s-j}) &= \text{Cov}(\mathbf{T}_{s-1}\boldsymbol{\alpha}_{s-1} + \mathbf{H}_{s-1}\boldsymbol{\varepsilon}_{s-1}, \mathbf{v}_{s-j}) \\
&= \mathbf{T}_{s-1}\text{Cov}(\boldsymbol{\alpha}_{s-1}, \mathbf{v}_{s-j}) \\
&= \mathbf{T}_{s-1}\text{Cov}(\boldsymbol{\alpha}_{s-1}, \mathbf{v}_{s-j}) \quad (\text{by induction}) \\
&= \mathbf{T}_{s-1}\text{Cov}(\boldsymbol{\alpha}_{s'}, \mathbf{v}_{s'-(j-1)}) \quad (\text{setting } s' = s - 1) \\
&= \mathbf{T}_{s-1}\mathbf{T}_{s'-(j-1)+1, s'-1}\mathbf{K}_{s'-(j-1)}\mathbf{F}_{s'-(j-1)} \\
&= \mathbf{T}_{s-1}\mathbf{T}_{s-j+1, s-2}\mathbf{K}_{s-j}\mathbf{F}_{s-j}
\end{aligned} \tag{4.57}$$

□

### 4.6.3 Filtering

#### State Estimate

Define the  $(r + 1)$ -step-ahead forecast  $\mathbf{a}_{s|s-r-1} = \mathbf{L}(\boldsymbol{\alpha}_s | Y_{s-r-1})$ . This is the MMSLE of the current state  $\boldsymbol{\alpha}_s$  given only the past of the series. It is equal to



$$\begin{aligned}
\mathbf{a}_{s|s-r-1} &= \mathbf{L}(\boldsymbol{\alpha}_s | \mathbf{Y}_{s-r-1}) \\
&= \mathbf{L}(\mathbf{T}_{s-1}\boldsymbol{\alpha}_{s-1} + \mathbf{H}_{s-1}\boldsymbol{\epsilon}_{s-1} | \mathbf{Y}_{s-r-1}) \\
&= \mathbf{T}_{s-1}\mathbf{L}(\boldsymbol{\alpha}_{s-1} | \mathbf{Y}_{s-r-1}) \\
&= \vdots \\
&= \mathbf{T}_{s-1} \dots \mathbf{T}_{s-r}\mathbf{L}(\boldsymbol{\alpha}_{s-r} | \mathbf{Y}_{s-r-1}) \\
&= \underline{\mathbf{T}}_{s-r,s-1}\mathbf{a}_{s-r}
\end{aligned} \tag{4.58}$$

Using linear estimation results from the technical introduction, we have

$$\begin{aligned}
\mathbf{a}_s &= \mathbf{L}(\boldsymbol{\alpha}_s | \mathbf{Y}_{s-1}) \\
&= \mathbf{L}(\boldsymbol{\alpha}_s | \mathbf{Y}_{s-r-1}, \mathbf{v}_{s-r}, \dots, \mathbf{v}_{s-1}) \\
&= \mathbf{L}(\boldsymbol{\alpha}_s | \mathbf{Y}_{s-r-1}) + \sum_{j=1}^r \text{Cov}(\boldsymbol{\alpha}_s, \mathbf{v}_{s-j}) \text{Var}(\mathbf{v}_{s-j})^{-1} \mathbf{v}_{s-j} \\
&= \underline{\mathbf{T}}_{s-r,s-1}\mathbf{a}_{s-r} + \sum_{j=1}^r \underline{\mathbf{T}}_{s-j+1,s-1} \mathbf{K}_{s-j} \mathbf{F}_{s-j} \mathbf{F}_{s-j}^{-1} \mathbf{v}_{s-j}
\end{aligned} \tag{4.59}$$

which can be written as

$$\mathbf{a}_s = \underbrace{\mathbf{a}_{s|s-r-1}}_{\text{past}} + \sum_{j=1}^r \underbrace{\underline{\mathbf{T}}_{s-j+1,s-1} \mathbf{K}_{s-j} \mathbf{v}_{s-j}}_{\text{contemporaneous}} \tag{4.60}$$

We have thus succeeded in expressing  $\mathbf{a}_s$  as a sum of the  $(r+1)$ -step-ahead state estimator  $\mathbf{a}_{s|s-r-1}$  and the contribution of the contemporaneous terms.

## State Estimate MSE

It is possible to separate the MSE of  $\mathbf{a}_s$  into past and present components in a similar way:

$$\begin{aligned}
& \text{MSE}_{\alpha_s}(\mathbf{a}_s) \\
&= \text{MSE}_{\alpha_s}[\mathbf{L}(\alpha_s | \mathbf{Y}_{s-r-1}, \mathbf{v}_{s-r}, \dots, \mathbf{v}_{s-1})] \\
&= \text{MSE}_{\alpha_s}[\mathbf{L}(\alpha_s | \mathbf{Y}_{s-r-1})] - \sum_{j=1}^r \text{Cov}(\alpha_s, \mathbf{v}_{s-j}) \text{Var}(\mathbf{v}_{s-j})^{-1} \text{Cov}(\mathbf{v}_{s-j}, \alpha_s) \\
&= \text{MSE}_{\alpha_s}(\mathbf{a}_{s|s-r-1}) - \sum_{j=1}^r \underline{\mathbf{T}}_{s-j+1, s-1} \mathbf{K}_{s-j} \mathbf{F}_{s-j} \mathbf{K}'_{s-j} \underline{\mathbf{T}}'_{s-j+1, s-1} \quad (4.61)
\end{aligned}$$

Let  $\mathbf{P}_{s|s-r-1}$  be the MSE of  $\mathbf{a}_{s|s-r-1}$  as a predictor of  $\alpha_s$ . It can be computed recursively:

$$\begin{aligned}
\mathbf{P}_{s|s-r-1} &= \text{Var}(\alpha_s - \mathbf{a}_{s|s-r-1}) \\
&= \text{Var}(\mathbf{T}_{s-1} \alpha_{s-1} + \mathbf{H}_{s-1} \epsilon_{s-1} - \mathbf{T}_{s-1} \mathbf{a}_{s-1|s-r-1}) \\
&= \mathbf{T}_{s-1} \text{Var}(\alpha_{s-1} - \mathbf{a}_{s-1|s-r-1}) \mathbf{T}'_{s-1} + \mathbf{H}_{s-1} \text{Var}(\epsilon_{s-1}) \mathbf{H}'_{s-1} \\
&= \mathbf{T}_{s-1} \mathbf{P}_{s-1|s-r-1} \mathbf{T}'_{s-1} + \mathbf{H}_{s-1} \mathbf{H}'_{s-1} \quad (4.62)
\end{aligned}$$

Substituting into (4.61) yields

$$\mathbf{P}_s = \underbrace{\mathbf{P}_{s|s-r-1}}_{\text{past}} - \sum_{j=1}^r \underbrace{\underline{\mathbf{T}}_{s-j+1, s-1} \mathbf{K}_{s-j} \mathbf{F}_{s-j} \mathbf{K}'_{s-j} \underline{\mathbf{T}}'_{s-j+1, s-1}}_{\text{contemporaneous}} \quad (4.63)$$

The summation term in this expression represents the reduction in MSE which results from taking into account the contemporaneous terms.

## Observation Forecast

The state estimate  $\mathbf{a}_s$  can be used to construct the corresponding one-step-ahead forecast of  $\mathbf{y}_s$ , which is  $\mathbf{y}_{s|s-1} \equiv \mathbf{L}(\mathbf{y}_s|\mathbf{Y}_{s-1})$ . It is easy to see that

$$\mathbf{y}_{s|s-1} = \mathbf{L}(\mathbf{Z}_s \boldsymbol{\alpha}_s + \mathbf{G}_s \boldsymbol{\varepsilon}_s | \mathbf{Y}_{s-1}) = \mathbf{Z}_s \mathbf{a}_s \quad (4.64)$$

since  $\boldsymbol{\varepsilon}_s$  is independent of  $\mathbf{Y}_{s-1}$ . The associated prediction error is the innovation  $\mathbf{v}_s$ , so the MSE is

$$\text{MSE}(\mathbf{y}_{s|s-1}) = \text{Var}(\mathbf{v}_s) = \mathbf{F}_s \quad (4.65)$$

Similarly to the state estimate, it is possible to separate the 1-step-ahead observation estimate  $\mathbf{y}_{s|s-1}$  and its MSE into past and contemporaneous components. The  $(r+1)$ -step-ahead predictor of  $\mathbf{y}_s$  is

$$\begin{aligned} \mathbf{y}_{s|s-r-1} &= \mathbf{L}(\mathbf{y}_s | \mathbf{Y}_{s-r-1}) \\ &= \mathbf{L}(\mathbf{Z}_s \boldsymbol{\alpha}_s + \mathbf{G}_s \boldsymbol{\varepsilon}_s | \mathbf{Y}_{s-r-1}) \\ &= \mathbf{Z}_s \mathbf{a}_{s|s-r-1} \\ &= \mathbf{Z}_s \mathbf{T}_{s-r,s-1} \mathbf{a}_{s-r} \end{aligned} \quad (4.66)$$

Using the decomposition (4.60) for  $\mathbf{a}_s$ , the 1-step-ahead predictor  $\mathbf{y}_{s|s-1}$  is

$$\begin{aligned} \mathbf{y}_{s|s-1} &= \mathbf{L}(\mathbf{y}_s | \mathbf{Y}_{s-1}) \\ &= \mathbf{Z}_s \mathbf{a}_s \\ &= \mathbf{Z}_s \mathbf{T}_{s-r,s-1} \mathbf{a}_{s-r} + \sum_{j=1}^r \mathbf{Z}_s \mathbf{T}_{s-j+1,s-1} \mathbf{K}_{s-j} \mathbf{v}_{s-j} \end{aligned} \quad (4.67)$$

## Observation Forecast MSE

The MSE of  $\mathbf{y}_{s|s-r-1}$  as an estimator of  $\mathbf{y}_s$  is

$$\begin{aligned}
\text{MSE}_{\mathbf{y}_s}(\mathbf{y}_{s|s-r-1}) &= \text{Var}(\mathbf{y}_s - \mathbf{y}_{s|s-r-1}) \\
&= \text{Var}(\mathbf{Z}_s \boldsymbol{\alpha}_s + \mathbf{G}_s \boldsymbol{\varepsilon}_s - \mathbf{Z}_s \mathbf{a}_{s|s-r-1}) \\
&= \mathbf{Z}_s \text{Var}(\boldsymbol{\alpha}_s - \mathbf{a}_{s|s-r-1}) \mathbf{Z}_s' + \mathbf{G}_s \text{Var}(\boldsymbol{\varepsilon}_s) \mathbf{G}_s' \\
&= \mathbf{Z}_s \mathbf{P}_{s|s-r-1} \mathbf{Z}_s' + \mathbf{G}_s \mathbf{G}_s' \tag{4.68}
\end{aligned}$$

Using decomposition (4.63) we can write

$$\begin{aligned}
&\text{MSE}_{\mathbf{y}_s}(\mathbf{y}_{s|s-1}) \\
&= \mathbf{Z}_s \mathbf{P}_s \mathbf{Z}_s' + \mathbf{G}_s \mathbf{G}_s' \\
&= \mathbf{Z}_s \mathbf{P}_{s|s-r-1} \mathbf{Z}_s' + \mathbf{G}_s \mathbf{G}_s' \\
&\quad - \sum_{j=1}^r \mathbf{Z}_s \mathbf{T}_{s-j+1,s-1} \mathbf{K}_{s-j} \mathbf{F}_{s-j} \mathbf{K}_{s-j}' \mathbf{T}_{s-j+1,s-1}' \mathbf{Z}_s' \\
&= \underbrace{\text{MSE}_{\mathbf{y}_s}(\mathbf{y}_{s|s-r-1})}_{\text{past}} - \sum_{j=1}^r \underbrace{\mathbf{Z}_s \mathbf{T}_{s-j+1,s-1} \mathbf{K}_{s-j} \mathbf{F}_{s-j} \mathbf{K}_{s-j}' \mathbf{T}_{s-j+1,s-1}' \mathbf{Z}_s'}_{\text{contemporaneous}} \\
&\tag{4.69}
\end{aligned}$$

As before, the summation term represents the reduction in the MSE that results from taking into account contemporaneous terms.

## 4.7 Conclusions

We have demonstrated how to construct the UR of any multivariate time series model that can be expressed in the general SSF. The technique of

incorporating the error term into the state vector allows us to sidestep the issue of correlated measurement and transition errors.

While the UR itself is not a novel concept, previous treatments have not considered the output from the UR KFS in its own right, but only as a means of obtaining the output of the MR KFS. We examined the theoretical properties of the output quantities, particularly the UR innovations, and illustrated how they can be used to construct estimators and deletion residuals that cannot be obtained directly from the MR KFS.

In Chapter 5 we continue this theoretical treatment by considering the UR in the context of the steady-state Kalman filter and examining the behaviour of the UR filter when the MR filter converges. Finally, several of the methods for estimation of missing temperature data in Chapter 6 rely heavily on the results established in this chapter, particularly the various deletion residuals obtained through the UR KFS.

## Chapter 5

# The Steady-State Filter and Periodic Convergence

### 5.1 Introduction

There are many situations where the system matrices in the SSF of a time series do not evolve over time. Some examples are the autoregressive models and the simple structural models discussed in the technical introduction. If the matrices  $\mathbf{Z}_t, \mathbf{T}_t, \mathbf{G}_t$  and  $\mathbf{H}_t$  do not depend on  $t$ , their subscripts are omitted and the model is said to be *time-invariant*. The updating equations in the KF for a time-invariant model often become redundant, as the output matrices converge to constant quantities (Harvey, 1989). This is an attractive property of the KF because it can drastically reduce the number of operations required at each recursive step.

In the first part of this chapter we examine the conditions on the system matrices for convergence to take place. The theoretical background of this topic is in control engineering (e.g. Caines & Mayne, 1970; Chan *et al.*, 1984) and applies to a different form of the state-space model, where there is assumed to be no measurement error (that is,  $\mathbf{G}_t = 0$ ). We discuss how this framework can be adapted to the general SSF of de Jong (1991) by using the technique of putting the error term in the state vector (see Chapter 4).

In the second part of the chapter, we consider the issues arising from

the fact that the UR of a time-invariant multivariate time series is, by construction, not time-invariant; for example, the UR transition matrix is equal to the MR transition matrix  $\mathbf{T}$  at the end of each time period  $(p_t, t)$ , and equal to  $\mathbf{I}$  otherwise. We prove that when the MR filter converges, the UR filter output matrices take a simple form and explain how it is possible to retain the computational savings of the steady-state KF even in this case. We then extend this approach to all state-space models where the system matrices vary periodically, and introduce the notion of *periodic convergence* of the KF.

We conclude by establishing a set of conditions on the system matrices of a periodic model, under which periodic convergence of the KF is achieved.

## 5.2 Filter Steady State

### 5.2.1 Time-invariant Models

A general SSF for a time series model with time-invariant system matrices is

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}\boldsymbol{\alpha}_t + \mathbf{G}\boldsymbol{\varepsilon}_t \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}\boldsymbol{\alpha}_t + \mathbf{H}\boldsymbol{\varepsilon}_t \quad , \quad t = 1, \dots, n \end{aligned} \tag{5.1}$$

where, as before,  $\{\boldsymbol{\varepsilon}_t\} \sim \text{WN}(\mathbf{0}, \sigma^2 \mathbf{I}_r)$  and  $\boldsymbol{\alpha}_1$  has mean  $\mathbf{a}_1$  and variance  $\mathbf{P}_1$ . The Kalman filter recursions for this model are

$$\begin{aligned}
\mathbf{v}_t &= \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t \\
\mathbf{F}_t &= \mathbf{Z} \mathbf{P}_t \mathbf{Z}' + \mathbf{G} \mathbf{G}' \\
\mathbf{K}_t &= (\mathbf{T} \mathbf{P}_t \mathbf{Z}' + \mathbf{H} \mathbf{G}') \mathbf{F}_t^{-1} \\
\mathbf{L}_t &= \mathbf{T} - \mathbf{K}_t \mathbf{Z} \\
\mathbf{a}_{t+1} &= \mathbf{T} \mathbf{a}_t + \mathbf{K}_t \mathbf{v}_t \\
\mathbf{P}_{t+1} &= \mathbf{T} \mathbf{P}_t \mathbf{L}_t' + \mathbf{H} (\mathbf{H} - \mathbf{K}_t \mathbf{G})'
\end{aligned} \tag{5.2}$$

It can be shown that, under certain conditions, the MSE of the 1-step-ahead state predictor converges to a constant matrix, that is,

$$\lim_{t \rightarrow \infty} \mathbf{P}_t = \bar{\mathbf{P}} \tag{5.3}$$

When this occurs, it is said that the KF is in a steady state. It is easy to see that, if  $\mathbf{P}_t \rightarrow \bar{\mathbf{P}}$ , then  $\mathbf{F}_t \rightarrow \bar{\mathbf{F}}$ ,  $\mathbf{K}_t \rightarrow \bar{\mathbf{K}}$  and  $\mathbf{L}_t \rightarrow \bar{\mathbf{L}}$  as  $t \rightarrow \infty$ , where

$$\begin{aligned}
\bar{\mathbf{F}} &= \mathbf{Z} \bar{\mathbf{P}} \mathbf{Z}' + \mathbf{G} \mathbf{G}' \\
\bar{\mathbf{K}} &= (\mathbf{T} \bar{\mathbf{P}} \mathbf{Z}' + \mathbf{H} \mathbf{G}') \bar{\mathbf{F}}^{-1} \\
\bar{\mathbf{L}} &= \mathbf{T} - \bar{\mathbf{K}} \mathbf{Z}
\end{aligned} \tag{5.4}$$

hence the KF output matrices are all time-invariant once  $\mathbf{P}_t$  converges to a constant matrix. This means that their recursions are redundant and only the quantities  $\mathbf{v}_t$  and  $\mathbf{a}_t$  need to be updated at each time period. In practice, we monitor the recursions until the difference  $\mathbf{P}_{t+1} - \mathbf{P}_t$  is small enough to deem the filter to have converged. Storing the steady-state matrices and omitting the corresponding recursions then leads to considerable computational savings (Harvey, 1989).



### 5.2.2 Definitions

The updating equation for the MSE of the 1-step-ahead state MMSLE can be written in the form

$$\begin{aligned}
 P_{t+1} &= TP_tT' - (TP_tZ' + HG'K'_t + HH') \\
 &= TP_tT' - TP_tZ' - HG'(ZP_tZ' + GG')^{-1}(TP_tZ' + HG')' - HH'
 \end{aligned} \tag{5.5}$$

The Kalman filter has a steady-state solution if there exists a time-invariant MSE matrix  $\bar{P}$  which is unchanged by the updating equation. If such a solution exists, equation (5.5) can be expressed in the form of an algebraic Riccati equation (ARE):

$$\bar{P} - T\bar{P}T' + T\bar{P}Z' + HG'(Z\bar{P}Z' + GG')^{-1}(T\bar{P}Z' + HG')' + HH' = 0 \tag{5.6}$$

In practice, the ARE is difficult to solve in all but the simplest models (e.g. Brogan, 1991). Despite this, it is possible to establish sufficient conditions for convergence by considering the properties of the system matrices.

We define the following matrix properties, where  $\mathbf{A}$  is a  $n \times n$  matrix,  $\mathbf{B}$  is a  $n \times m$  matrix and  $\mathbf{C}$  is a  $m \times n$  matrix:

- The matrix  $\mathbf{A}$  is *stable* if  $|\lambda_i(\mathbf{A})| < 1$  for  $i = 1, \dots, n$ ; that is, all of its eigenvalues lie within the unit circle.
- The matrix system  $(\mathbf{A}, \mathbf{B})$  is *controllable* if the rows of the  $n \times mn$  control matrix  $[\mathbf{B}, \mathbf{A}\mathbf{B}, \dots, \mathbf{A}^{n-1}\mathbf{B}]$  are linearly independent. We can also use the equivalent definition:  $(\mathbf{A}, \mathbf{B})$  is controllable if  $\forall i \in \{1, \dots, n\}$  such that  $\lambda_i(\mathbf{A})' \neq 0$ , we have  $\mathbf{B}'\mathbf{v}(\lambda_i) \neq 0$ , where

$\mathbf{v}(\lambda_i)$  is the eigenvector corresponding to  $\lambda_i$ ; that is, when all eigenvectors of  $\mathbf{A}'$  corresponding to non-zero eigenvalues lie outside the kernel of  $\mathbf{B}'$ .

- The matrix system  $(\mathbf{A}, \mathbf{B})$  is *stabilisable* if there exists a  $m \times n$  matrix  $\mathbf{S}$  such that  $|\lambda_i(\mathbf{A} + \mathbf{BS})| < 1$  for  $i = 1, \dots, n$ . An equivalent definition is:  $(\mathbf{A}, \mathbf{B})$  is stabilisable if  $\forall i \in \{1, \dots, n\}$  such that  $|\lambda_i(\mathbf{A})| \geq 1$ , we have  $\mathbf{B}'\mathbf{v}(\lambda_i) \neq 0$ ; that is, when all eigenvectors of  $\mathbf{A}'$  corresponding to eigenvalues outside the unit circle lie outside the kernel of  $\mathbf{B}'$ .
- The matrix system  $(\mathbf{A}, \mathbf{C})$  is *observable* if the rows of the  $n \times mn$  observation matrix  $[\mathbf{C}', (\mathbf{CA})', \dots, (\mathbf{CA}^{n-1})']$  are linearly independent. Equivalently:  $(\mathbf{A}, \mathbf{C})$  is observable if  $\forall i \in \{1, \dots, n\}$  such that  $\lambda_i(\mathbf{A}) \neq 0$ , we have  $\mathbf{C}\mathbf{v}(\lambda_i) \neq 0$ .
- The matrix system  $(\mathbf{A}, \mathbf{C})$  is *detectable* if there exists a  $n \times m$  matrix  $\mathbf{D}$  such that  $|\lambda_i(\mathbf{A} + \mathbf{DC})| < 1$  for  $i = 1, \dots, n$ . Equivalently:  $(\mathbf{A}, \mathbf{C})$  is detectable if  $\forall i \in \{1, \dots, n\}$  such that  $|\lambda_i(\mathbf{A})| \geq 1$ , we have  $\mathbf{C}\mathbf{v}(\lambda_i) \neq 0$ .

From these definitions, we can deduce the following:

1. Controllability implies stabilisability, and observability implies detectability, but not vice versa. A sufficient, but not necessary condition for the converse to hold is  $\det(\mathbf{A}) \neq 0$ . In this case,  $\lambda = 0$  is not an eigenvalue of  $\mathbf{A}$ , so the conditions are equivalent.
2. If the rows of  $\mathbf{B}$  are linearly independent, then, for any choice of  $\mathbf{A}$ , the rows of the control matrix are also linearly independent, hence  $(\mathbf{A}, \mathbf{B})$  is controllable. Similarly, if the columns of  $\mathbf{C}$  are linearly independent, then, for any choice of  $\mathbf{A}$ , the rows of the observation matrix are also linearly independent, hence  $(\mathbf{A}, \mathbf{C})$  is observable.

3. If  $\mathbf{A}$  is stable, then it is both stabilisable and detectable; setting  $\mathbf{D} = \mathbf{0}$  and  $\mathbf{S} = \mathbf{0}$  is enough to satisfy the conditions.
4. Stabilisability and detectability are dual properties; if  $(\mathbf{A}, \mathbf{B})$  is stabilisable then  $(\mathbf{A}', \mathbf{B}')$  is detectable, and vice versa. This can be verified by setting  $\mathbf{D} = \mathbf{S}'$ . Similarly, the properties of controllability and observability are also dual.

### 5.2.3 Conditions for Convergence

We now consider the model

$$\begin{aligned}\mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t\end{aligned}\tag{5.7}$$

This representation has its basis in control theory and does not include an error term in the measurement equation. Also, the interpretation of the model is different: the aim here is to choose appropriate values for the *control variables*  $\{\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_i\}$  to steer the model towards a particular state  $\boldsymbol{\alpha}_{i+1}$ , with the added difficulty that the state is observed only through its effect on the measurements  $\{\mathbf{y}_t\}$ .

Based on the properties of the system matrices of (5.7), we define the following properties:

**Stability** : The model is *stable* if the matrix  $\mathbf{T}$  is stable.

**Controllability** : The model is *controllable* if  $(\mathbf{T}, \mathbf{H})$  is controllable.

**Observability** : The model is *observable* if  $(\mathbf{T}, \mathbf{Z})$  is observable.

**Stabilisability** : The model is *stabilisable* if  $(\mathbf{T}, \mathbf{H})$  is stabilisable.

**Detectability** : The model is *detectable* if  $(\mathbf{T}, \mathbf{Z})$  is detectable.

Caines & Mayne (1970) prove that, if the initial variance matrix  $\mathbf{P}_1$

is positive semi-definite and the model is stable then the MSE matrix  $\mathbf{P}_t$  converges to  $\bar{\mathbf{P}}$  exponentially fast, as long as  $\bar{\mathbf{P}}$  is the only positive semi-definite solution to the ARE. Anderson & Moore (1979) show that this also holds if the system is detectable and stabilisable, but not necessarily stable. This is a more general result, because stability implies detectability and stabilisability, but not vice versa. Chan *et al.* (1984) establish that if the system is observable and if  $\mathbf{P}_1 - \bar{\mathbf{P}}$  is positive definite or zero, then  $\mathbf{P}_t$  still converges to  $\bar{\mathbf{P}}$ , but not necessarily exponentially fast.

#### 5.2.4 General SSF

We now examine how the properties for model (5.7) translate into properties for the more general state-space model. Recall that the model

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{G}_t \boldsymbol{\varepsilon}_t \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \boldsymbol{\varepsilon}_t \end{aligned} \quad (5.8)$$

can be written in the form

$$\begin{aligned} \mathbf{y}_t &= \begin{bmatrix} \mathbf{Z}_t & \mathbf{G}_t \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_t \\ \boldsymbol{\varepsilon}_t \end{bmatrix} \\ \begin{bmatrix} \boldsymbol{\alpha}_{t+1} \\ \boldsymbol{\varepsilon}_{t+1} \end{bmatrix} &= \begin{bmatrix} \mathbf{T}_t & \mathbf{H}_t \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_t \\ \boldsymbol{\varepsilon}_t \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \boldsymbol{\varepsilon}_{t+1} \end{aligned} \quad (5.9)$$

and let  $\widehat{\mathbf{Z}}_t, \widehat{\mathbf{T}}_t, \widehat{\mathbf{H}}_t$  (and  $\widehat{\mathbf{G}}_t = \mathbf{0}$ ) be the system matrices in this representation. The characteristic polynomial of the transition matrix is

$$\begin{aligned}
\det(\widehat{\mathbf{T}}_t - \lambda \mathbf{I}) &= \begin{vmatrix} \mathbf{T}_t - \lambda \mathbf{I} & \mathbf{H}_t \\ \mathbf{0} & -\lambda \mathbf{I} \end{vmatrix} \\
&= -\lambda \det(\mathbf{T}_t - \lambda \mathbf{I})
\end{aligned} \tag{5.10}$$

Its roots are the eigenvalues of  $\mathbf{T}_t$  and zero, thus  $\widehat{\mathbf{T}}_t$  is stable iff  $\mathbf{T}_t$  is stable.

Now let  $\mathbf{v}(\lambda) = [\mathbf{v}'_1 \mid \mathbf{v}'_2]'$  be an eigenvector of  $\widehat{\mathbf{T}}_t$ , for  $\lambda \neq 0$ . The vector  $\mathbf{v}(\lambda)$  is partitioned conformally to the block matrix  $\widehat{\mathbf{T}}_t$ . By definition:

$$\begin{aligned}
&\det(\widehat{\mathbf{T}}_t - \lambda \mathbf{I})\mathbf{v}(\lambda) = \mathbf{0} \\
&\Leftrightarrow \begin{bmatrix} \mathbf{T}_t - \lambda \mathbf{I} & \mathbf{H}_t \\ \mathbf{0} & -\lambda \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} = \mathbf{0} \\
&\Leftrightarrow \begin{bmatrix} (\mathbf{T}_t - \lambda \mathbf{I})\mathbf{v}_1 + \mathbf{H}_t\mathbf{v}_2 \\ -\lambda\mathbf{v}_2 \end{bmatrix} = \mathbf{0} \\
&\Rightarrow \mathbf{v}(\lambda) = \begin{bmatrix} \mathbf{v}_1(\lambda) \\ \mathbf{0} \end{bmatrix}
\end{aligned} \tag{5.11}$$

where  $\mathbf{v}_1(\lambda)$  is an eigenvector of  $\mathbf{T}_t$ , corresponding to the same eigenvalue  $\lambda$ . Consider the expression

$$\begin{aligned}
\widehat{\mathbf{Z}}_t \mathbf{v}(\lambda) &= \begin{bmatrix} \mathbf{Z}_t & \mathbf{G}_t \end{bmatrix} \begin{bmatrix} \mathbf{v}_1(\lambda) \\ \mathbf{0} \end{bmatrix} \\
&= \mathbf{Z}_t \mathbf{v}_1(\lambda)
\end{aligned} \tag{5.12}$$

This is non-zero iff  $\mathbf{Z}_t \mathbf{v}_1(\lambda)$  is non-zero, so we can deduce that  $(\widehat{\mathbf{T}}_t, \widehat{\mathbf{Z}}_t)$  is detectable (observable) iff  $(\mathbf{T}_t, \mathbf{Z}_t)$  is detectable (observable).

Finally, let  $\mathbf{w}(\lambda) = \begin{bmatrix} \mathbf{w}'_1 & \mathbf{w}'_2 \end{bmatrix}'$  be an eigenvector of  $\widehat{\mathbf{T}}'_t$ , for  $\lambda \neq 0$ . We have:

$$\begin{aligned}
&\det(\widehat{\mathbf{T}}'_t - \lambda \mathbf{I}) \mathbf{w}(\lambda) = \mathbf{0} \\
&\Leftrightarrow \begin{bmatrix} \mathbf{T}'_t - \lambda \mathbf{I} & \mathbf{0} \\ \mathbf{H}'_t & -\lambda \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} = \mathbf{0} \\
&\Leftrightarrow \begin{bmatrix} (\mathbf{T}'_t - \lambda \mathbf{I}) \mathbf{w}_1 \\ \mathbf{H}'_t \mathbf{w}_1 - \lambda \mathbf{w}_2 \end{bmatrix} = \mathbf{0} \\
&\Rightarrow \mathbf{w}(\lambda) = \begin{bmatrix} \mathbf{w}_1(\lambda) \\ \frac{1}{\lambda} \mathbf{H}'_t \mathbf{w}_1(\lambda) \end{bmatrix}
\end{aligned} \tag{5.13}$$

where  $\mathbf{w}_1(\lambda)$  is an eigenvector of  $\mathbf{T}'_t$ , corresponding to the same eigenvalue  $\lambda$ . Consider the expression

$$\begin{aligned}
\widehat{\mathbf{H}}'_t \mathbf{w}(\lambda) &= \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{w}_1(\lambda) \\ \frac{1}{\lambda} \mathbf{H}'_t \mathbf{w}_1(\lambda) \end{bmatrix} \\
&= \frac{1}{\lambda} \mathbf{H}'_t \mathbf{w}_1(\lambda)
\end{aligned} \tag{5.14}$$

This is non-zero iff  $\mathbf{H}_t' \mathbf{w}_1(\lambda)$  is non-zero, so we can deduce that  $(\hat{\mathbf{T}}_t, \hat{\mathbf{H}}_t)$  is stabilisable (controllable) iff  $(\mathbf{T}_t, \mathbf{H}_t)$  is stabilisable (controllable).

Thus, it is not necessary to eliminate the measurement error. The definitions of all the properties we examined can be applied directly to the general state-space model.

### 5.2.5 Application

We will now examine how convergence of the MR KF affects the UR filter recursions. It is clear that the regular definitions of convergence do not apply, since the UR does not have time-invariant system matrices. For example, the transition matrix  $\mathbf{T}_{i,t}$  is equal to  $\mathbf{T}_t$  when  $i = p$  and  $\mathbf{I}_q$  otherwise. However, since the system matrices vary in a very simple way, it is possible to adapt the definition of convergence to this situation. Note that we have to impose the additional condition  $p_t = p$ . A multivariate model where the size of the observation vector varies over time cannot have time-invariant system matrices.

The UR of a time-invariant model is

$$\begin{aligned} y_{i,t} &= \tilde{\mathbf{Z}}_i \boldsymbol{\alpha}_{i,t} \\ \boldsymbol{\alpha}_{i,t+1} &= \begin{cases} \mathbf{T} \boldsymbol{\alpha}_{p,t} + \mathbf{H} \boldsymbol{\epsilon}_t & \text{if } i = 1 \\ \boldsymbol{\alpha}_{i-1,t+1} & \text{if } i = 2, 3, \dots, p \end{cases} \end{aligned} \quad (5.15)$$

where  $\tilde{\mathbf{Z}}_i$  is the  $i^{th}$  row of  $\mathbf{Z}$ . Assume that the multivariate filter converges. We have already established that  $\tilde{\mathbf{P}}_{i,t}$  is equal to  $\mathbf{P}_t$  when  $i=1$ , since both matrices are equal to the MSE of the MMSLE of  $\boldsymbol{\alpha}_t$  given  $\mathbf{Y}_{t-1}$ . Now consider the recursive step at time  $(1,t)$ :

$$\begin{aligned}
\tilde{v}_{1,t} &= y_{1,t} - \tilde{Z}_1 \tilde{a}_{1,t} \\
\tilde{F}_{1,t} &= \tilde{Z}_1 \bar{P} \tilde{Z}_1' \equiv \bar{F}_1 \\
\tilde{K}_{1,t} &= T \bar{P} \tilde{Z}_1' \bar{F}_1^{-1} \equiv \bar{K}_1 \\
\tilde{a}_{2,t} &= \tilde{a}_{1,t} + \bar{K}_1 \tilde{v}_{1,t} \\
\tilde{P}_{2,t} &= \bar{P} - \bar{P} \tilde{Z}_1' \bar{K}_1' \equiv \bar{P}_2
\end{aligned} \tag{5.16}$$

We use the notation  $\bar{F}_1, \bar{K}_1, \bar{P}_2$  to indicate that these matrices do not depend on  $t$ . Repeating the step again, we find  $\bar{F}_2, \bar{K}_2, \bar{P}_3$ . These are also time-invariant; they only depend on  $i$ , the current position in the multivariate vector. Repeating the recursive steps at time  $t+1$  will only yield the same sequence of  $\bar{F}_i, \bar{K}_i, \bar{P}_i$ . This implies that the UR filter recursions converge periodically in the sense that

$$\begin{aligned}
\lim_{t \rightarrow \infty} \tilde{F}_{i,t} &= \bar{F}_i \\
\lim_{t \rightarrow \infty} \tilde{K}_{i,t} &= \bar{K}_i \\
\lim_{t \rightarrow \infty} \tilde{P}_{i,t} &= \bar{P}_i
\end{aligned} \tag{5.17}$$

for  $i = 1, \dots, p-1$ . Thus, it is far more efficient computationally to store these matrices once the KF converges, rather than calculate them at each pass. An important advantage to this approach is that the calculation of partially-interpolated estimators will involve much fewer operations.

For example, constructing the “past-and-present” estimator  $L(y_{i,t} | \mathbf{Y}_{t-1}, \mathbf{y}_{t \setminus i})$  involves evaluating the matrix product  $\tilde{L}_{i+1,t}' \dots \tilde{L}_{p_t,t}'$ , where  $\tilde{L}_{i,t} = \mathbf{T}_{i,t} - \tilde{K}_{i,t} \mathbf{Z}_{i,t}$ , the rest of the terms being readily available from the KF and KFS output. If the UR filter converges periodically, the matrix products become

$$\tilde{L}_{i+1,t}' \dots \tilde{L}_{p_t,t}' = \bar{L}_{i+1}' \dots \bar{L}_p' \equiv \bar{L}_{i+1|p}' \tag{5.18}$$



The terms  $\bar{\mathbf{L}}_{i|p}$ ,  $i = 1, \dots, p$ , do not depend on  $t$ . Thus we can store them once the filter converges and construct the estimates directly from the KFS output.

## 5.3 Generalisation to Periodic Models

### 5.3.1 Introduction

In the previous section we exploited the periodic structure of the UR SSF to simplify the KF recursions if the MR filter converges. It is possible to generalise this method to all other time series models where the system matrices are periodic. Consider a model where the system matrices  $\mathbf{Z}_t, \mathbf{T}_t, \mathbf{G}_t, \mathbf{H}_t$  contain periodic elements, with period  $s$ . We can use double-subscript notation to make the periodic structure explicit:

$$\begin{aligned} \mathbf{y}_{i,t} &= \mathbf{Z}_i \boldsymbol{\alpha}_{i,t} + \mathbf{G}_i \boldsymbol{\varepsilon}_{i,t} \\ \boldsymbol{\alpha}_{i,t+1} &= \begin{cases} \mathbf{T}_s \boldsymbol{\alpha}_{s,t} + \mathbf{H}_s \boldsymbol{\varepsilon}_{s,t} & \text{if } i = 1 \\ \mathbf{T}_{i-1} \boldsymbol{\alpha}_{i-1,t+1} + \mathbf{H}_{i-1} \boldsymbol{\varepsilon}_{i-1,t+1} & \text{if } i = 2, \dots, s \end{cases} \end{aligned} \quad (5.19)$$

for  $i = 1, \dots, s$  and  $t = 1, \dots, n$ , where  $\boldsymbol{\alpha}_{1,1} \sim (\mathbf{a}_{1,1}, \mathbf{P}_{1,1})$ . The UR can be thought of as a special case of this model, with  $\mathbf{G}_i = \mathbf{0}$  for  $i = 1, \dots, s$  and  $\mathbf{T}_i = \mathbf{I}$ ,  $\mathbf{H}_i = \mathbf{0}$  for  $i = 1, \dots, s-1$ . The KF recursions for the periodic model are

$$\mathbf{v}_{i,t} = \mathbf{y}_{i,t} - \mathbf{Z}_i \mathbf{a}_{i,t}$$

$$\mathbf{F}_{i,t} = \mathbf{Z}_i \mathbf{P}_{i,t} \mathbf{Z}_i' + \mathbf{G}_i \mathbf{G}_i'$$

$$\mathbf{K}_{i,t} = (\mathbf{T}_i \mathbf{P}_{i,t} \mathbf{Z}_i' + \mathbf{G}_i \mathbf{H}_i') \mathbf{F}_{i,t}^{-1}$$

$$\mathbf{L}_{i,t} = \mathbf{T}_i - \mathbf{K}_{i,t} \mathbf{Z}_i$$

$$\mathbf{a}_{i,t+1} = \begin{cases} \mathbf{T}_s \mathbf{a}_{s,t} + \mathbf{K}_{s,t} \mathbf{v}_{s,t} & \text{if } i = 1 \\ \mathbf{T}_{i-1} \mathbf{a}_{i-1,t+1} + \mathbf{K}_{i-1,t+1} \mathbf{v}_{i-1,t+1} & \text{if } i = 2, \dots, s \end{cases}$$

$$\mathbf{P}_{i,t+1} = \begin{cases} \mathbf{T}_s \mathbf{P}_{s,t} \mathbf{L}_{s,t} + \mathbf{H}_{s,t} (\mathbf{H}_{s,t} - \mathbf{K}_{s,t} \mathbf{G}_{s,t})' & \text{if } i = 1 \\ \mathbf{T}_{i-1} \mathbf{P}_{i-1,t+1} \mathbf{L}_{i-1,t+1} + \mathbf{H}_{i-1,t+1} (\mathbf{H}_{i-1,t+1} - \mathbf{K}_{i-1,t+1} \mathbf{G}_{i-1,t+1})' & \text{if } i = 2, \dots, s \end{cases} \quad (5.20)$$

In the exceptional case where the periodic elements of the system matrices do not have the same period, we can define  $s$  to be the lowest common multiple of the different periods. This situation could arise, for example, if  $\mathbf{y}_{i,t}$  are hourly observations and  $\mathbf{Z}_i$  has hourly seasonality, while  $\mathbf{T}_i$  has daily seasonality.

### 5.3.2 Periodic Convergence

As in the UR, the system matrices are not constant, so the Kalman filter will not converge to a steady state in the strict sense. However, it is possible that it will converge periodically, in the sense that

$$\lim_{t \rightarrow \infty} \mathbf{P}_{i,t} = \bar{\mathbf{P}}_i \quad (5.21)$$

for  $i = 1, \dots, p$ . In direct analogy to the way the scalar observations of the UR are stacked to form the MR, we can stack the (vector) ob-

servations of the periodic model to form a time-invariant state-space model:

$$\begin{bmatrix} \mathbf{y}_{1,t} \\ \vdots \\ \mathbf{y}_{s,t} \end{bmatrix} = \widehat{\mathbf{Z}} \begin{bmatrix} \boldsymbol{\alpha}_{1,t} \\ \vdots \\ \boldsymbol{\alpha}_{s,t} \\ \text{---} \text{---} \text{---} \\ \boldsymbol{\varepsilon}_{1,t} \\ \vdots \\ \boldsymbol{\varepsilon}_{s,t} \end{bmatrix}$$

$$\begin{bmatrix} \boldsymbol{\alpha}_{1,t+1} \\ \vdots \\ \boldsymbol{\alpha}_{s,t+1} \\ \text{---} \text{---} \text{---} \\ \boldsymbol{\varepsilon}_{1,t+1} \\ \vdots \\ \boldsymbol{\varepsilon}_{s,t+1} \end{bmatrix} = \widehat{\mathbf{T}} \begin{bmatrix} \boldsymbol{\alpha}_{1,t} \\ \vdots \\ \boldsymbol{\alpha}_{s,t} \\ \text{---} \text{---} \text{---} \\ \boldsymbol{\varepsilon}_{1,t} \\ \vdots \\ \boldsymbol{\varepsilon}_{s,t} \end{bmatrix} + \widehat{\mathbf{H}} \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t} \\ \vdots \\ \boldsymbol{\varepsilon}_{s,t} \end{bmatrix} \quad (5.22)$$

where we define the following system matrices:

$$\widehat{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z}_1 & & & \vdots & \mathbf{G}_1 & & \\ & \ddots & & & & \ddots & \\ & & \mathbf{Z}_s & \vdots & & & \mathbf{G}_s \end{bmatrix} \quad (5.23)$$

$$\widehat{\mathbf{T}} = \left[ \begin{array}{cc|cc} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}_1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_{s-1} \cdots \mathbf{T}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_{s-1} \cdots \mathbf{T}_2 \mathbf{H}_1 \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_s \cdots \mathbf{T}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_s \cdots \mathbf{T}_2 \mathbf{H}_1 \\ \hline \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \vdots & \ddots & & \vdots & \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \end{array} \right] \quad (5.24)$$

$$\widehat{\mathbf{H}} = \left[ \begin{array}{cccc|cccc} \mathbf{0} & & \cdots & \cdots & \cdots & \mathbf{0} & & \\ \hline \mathbf{H}_2 & & & & & \mathbf{0} & & \\ \mathbf{T}_3 \mathbf{H}_2 & & \mathbf{H}_3 & & & \mathbf{0} & & \\ \vdots & & \vdots & & \ddots & \vdots & & \\ \mathbf{T}_s \cdots \mathbf{T}_3 \mathbf{H}_2 & \mathbf{T}_s \cdots \mathbf{T}_4 \mathbf{H}_3 & \cdots & \mathbf{H}_s & \mathbf{0} & & & \\ \hline \mathbf{I} & & & & & & & \\ & & \mathbf{I} & & & & & \\ & & & \ddots & & & & \\ & & & & \mathbf{I} & & & \end{array} \right] \quad (5.25)$$

The time-invariant model has longer state and observation vectors, so filtering and smoothing will involve inverting larger matrices. For this reason we prefer to work directly with the periodic model in practice. However, the structure of the time-invariant model (constant system matrices, no measurement error) allows us to establish conditions for periodic convergence which can be expressed in terms of the system matrices of the periodic model.

We begin by proving that steady-state convergence in the time-invariant model is equivalent to periodic convergence in the periodic model. The

KF recursions for the time-invariant model will output the 1-step-ahead MMSLE of the state  $\hat{\alpha}_t$ . This is

$$\begin{aligned}\hat{\mathbf{a}}_t &= \mathbf{E}(\hat{\alpha}_t | \hat{\mathbf{Y}}_{t-1}) \\ &= \begin{bmatrix} \mathbf{E}(\alpha_{1,t} | \mathbf{Y}_{s,t-1}) \\ \vdots \\ \mathbf{E}(\alpha_{s,t} | \mathbf{Y}_{s,t-1}) \\ \text{-----} \\ \mathbf{E}(\epsilon_{1,t} | \mathbf{Y}_{s,t-1}) \\ \vdots \\ \mathbf{E}(\epsilon_{s,t} | \mathbf{Y}_{s,t-1}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{a}_{1,t|t-1} \\ \vdots \\ \mathbf{a}_{s,t|t-1} \\ \text{-----} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad (5.26)\end{aligned}$$

where  $\mathbf{a}_{i,t|t-1}$  is the  $i$ -step-ahead predictor of  $\alpha_{i,t}$ , and we write  $\hat{\mathbf{Y}}_{t-1} = \mathbf{Y}_{s,t-1}$  for consistency of notation. The KF recursions will also produce the MSE of  $\mathbf{a}_{i,t|t-1}$  as an estimator of  $\alpha_{i,t}$ . Define the  $i^{th}$  component of estimation error as  $\Delta \mathbf{a}_{1,t|t-1} \equiv \mathbf{a}_{1,t|t-1} - \alpha_{1,t}$ . We can express the MSE matrix as

$$\widehat{\mathbf{P}}_t = \text{MSE}(\widehat{\mathbf{a}}_t | \widehat{\boldsymbol{\alpha}}_t)$$

$$= \text{E}[(\widehat{\mathbf{a}}_t - \widehat{\boldsymbol{\alpha}}_t)(\widehat{\mathbf{a}}_t - \widehat{\boldsymbol{\alpha}}_t)']$$

$$= \begin{bmatrix} \text{E}(\Delta \mathbf{a}_{1,t|t-1} \Delta \mathbf{a}'_{1,t|t-1}) & \cdots & \text{E}(\Delta \mathbf{a}_{1,t|t-1} \Delta \mathbf{a}'_{s,t|t-1}) \\ \vdots & \ddots & \vdots \\ \text{E}(\Delta \mathbf{a}_{s,t|t-1} \Delta \mathbf{a}'_{1,t|t-1}) & \cdots & \text{E}(\Delta \mathbf{a}_{s,t|t-1} \Delta \mathbf{a}'_{s,t|t-1}) \\ \hline \text{E}[(0 - \boldsymbol{\varepsilon}_{1,t}) \Delta \mathbf{a}'_{1,t|t-1}] & \cdots & \text{E}[(0 - \boldsymbol{\varepsilon}_{1,t}) \Delta \mathbf{a}'_{s,t|t-1}] \\ \vdots & \ddots & \vdots \\ \text{E}[(0 - \boldsymbol{\varepsilon}_{s,t}) \Delta \mathbf{a}'_{1,t|t-1}] & \cdots & \text{E}[(0 - \boldsymbol{\varepsilon}_{s,t}) \Delta \mathbf{a}'_{s,t|t-1}] \end{bmatrix}$$

$$= \begin{bmatrix} \text{E}[\Delta \mathbf{a}_{1,t|t-1} (0 - \boldsymbol{\varepsilon}_{1,t})'] & \cdots & \text{E}[\Delta \mathbf{a}_{1,t|t-1} (0 - \boldsymbol{\varepsilon}_{s,t})'] \\ \vdots & \ddots & \vdots \\ \text{E}[\Delta \mathbf{a}_{s,t|t-1} (0 - \boldsymbol{\varepsilon}_{1,t})'] & \cdots & \text{E}[\Delta \mathbf{a}_{s,t|t-1} (0 - \boldsymbol{\varepsilon}_{s,t})'] \\ \hline \text{E}[(0 - \boldsymbol{\varepsilon}_{1,t})(0 - \boldsymbol{\varepsilon}_{1,t})'] & \cdots & \text{E}[(0 - \boldsymbol{\varepsilon}_{1,t})(0 - \boldsymbol{\varepsilon}_{s,t})'] \\ \vdots & \ddots & \vdots \\ \text{E}[(0 - \boldsymbol{\varepsilon}_{s,t})(0 - \boldsymbol{\varepsilon}_{1,t})'] & \cdots & \text{E}[(0 - \boldsymbol{\varepsilon}_{s,t})(0 - \boldsymbol{\varepsilon}_{s,t})'] \end{bmatrix} \quad (5.27)$$

The top-left-hand block of  $\widehat{\mathbf{P}}_t$  is equal to

$$\begin{aligned} \text{E}[(\mathbf{a}_{1,t|t-1} - \boldsymbol{\alpha}_{1,t})(\mathbf{a}_{1,t|t-1} - \boldsymbol{\alpha}_{1,t})'] &= \text{Var}(\mathbf{a}_{1,t|t-1} - \boldsymbol{\alpha}_{1,t}) \\ &= \text{MSE}(\mathbf{a}_{1,t|t-1} | \boldsymbol{\alpha}_{1,t}) \\ &= \mathbf{P}_{1,t} \end{aligned} \quad (5.28)$$

If  $\widehat{\mathbf{P}}_t$  converges to a steady-state solution, then all of its component

blocks must also converge. Thus,  $\mathbf{P}_{1,t} \rightarrow \bar{\mathbf{P}}_1$  as  $t \rightarrow \infty$ . Using the same argument as in the previous section, the fact that  $\mathbf{P}_{1,t}$  converges implies that  $\mathbf{P}_{i,t} \rightarrow \bar{\mathbf{P}}_i$  as  $t \rightarrow \infty$ , for all the other values of  $i$ . Hence the periodic KF will converge periodically.

To prove the converse, we begin by assuming periodic convergence of the periodic KF. It suffices to show that all of the blocks that comprise  $\hat{\mathbf{P}}_t$  will converge. Using results from the previous section:

$$\mathbb{E}[(0 - \varepsilon_{i,t})(0 - \varepsilon_{j,t})'] = \begin{cases} \mathbf{I} & \text{if } i = j \\ \mathbf{0} & \text{if } i \neq j \end{cases} \quad (5.29)$$

$$\mathbb{E}[(\mathbf{a}_{i,t|t-1} - \boldsymbol{\alpha}_{i,t})(0 - \varepsilon_{j,t})'] = \begin{cases} \mathbf{T}_{i-1} \dots \mathbf{T}_{j+1} \mathbf{H}_j & \text{if } i > j \\ \mathbf{0} & \text{if } i \leq j \end{cases} \quad (5.30)$$

which are all independent of  $t$ . To find the remaining terms, assume, without loss of generality,  $i \geq j$ . We have

$$\begin{aligned} & \mathbb{E}[(\mathbf{a}_{i,t|t-1} - \boldsymbol{\alpha}_{i,t})(\mathbf{a}_{j,t|t-1} - \boldsymbol{\alpha}_{j,t})'] \\ &= \mathbf{T}_{i-1} \dots \mathbf{T}_j \mathbb{E}[(\mathbf{a}_{j,t|t-1} - \boldsymbol{\alpha}_{j,t})(\mathbf{a}_{j,t|t-1} - \boldsymbol{\alpha}_{j,t})'] \\ &= \mathbf{T}_{i-1} \dots \mathbf{T}_j \text{MSE}_{\boldsymbol{\alpha}_{j,t}}(\mathbf{a}_{j,t|t-1}) \end{aligned} \quad (5.31)$$

We established previously that the MSE of the  $j$ -step-ahead predictor of  $\boldsymbol{\alpha}_{j,t}$  depends only on the system matrices and  $\mathbf{P}_{1,t}$ . The system matrices do not depend on  $t$  and we have assumed that  $\mathbf{P}_{1,t} \rightarrow \bar{\mathbf{P}}_1$ . Thus, as  $t \rightarrow \infty$  none of the elements of  $\hat{\mathbf{P}}_t$  depend on  $t$  and the KF converges to a steady state.

### 5.3.3 Stability

We will first examine the conditions under which the matrix  $\widehat{\mathbf{T}}$  is stable.

The characteristic polynomial is

$$\begin{aligned}
 & \det(\widehat{\mathbf{T}} - \lambda \mathbf{I}) \\
 &= \begin{vmatrix} -\lambda \mathbf{I} & \cdots & \mathbf{0} & \mathbf{T}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}_1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & -\lambda \mathbf{I} & \mathbf{T}_{s-1} \dots \mathbf{T}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_{s-1} \dots \mathbf{T}_2 \mathbf{H}_1 \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_s \dots \mathbf{T}_1 - \lambda \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_s \dots \mathbf{T}_2 \mathbf{H}_1 \\ \hline \mathbf{0} & \cdots & \cdots & \mathbf{0} & -\lambda \mathbf{I} & \cdots & \cdots & \mathbf{0} \\ \vdots & \ddots & & \vdots & \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \cdots & -\lambda \mathbf{I} \end{vmatrix} \\
 &= (-\lambda)^{2s-1} \det(\mathbf{T}_s \dots \mathbf{T}_1 - \lambda \mathbf{I}) \\
 &= (-\lambda)^{2s-1} \det(\underline{\mathbf{T}} - \lambda \mathbf{I}) \tag{5.32}
 \end{aligned}$$

where we define the summary matrix  $\underline{\mathbf{T}} = \mathbf{T}_s \dots \mathbf{T}_1$ . We can deduce that the eigenvalues of  $\widehat{\mathbf{T}}$  are the eigenvalues of  $\underline{\mathbf{T}}$  and zero. Thus, the matrix  $\widehat{\mathbf{T}}$  is stable if and only if  $\underline{\mathbf{T}}$  is stable.

An interesting corollary of this result arises when the transition matrix is time-invariant, that is  $\mathbf{T}_1 = \dots = \mathbf{T}_s = \mathbf{T}$ . There are many situations where such a model is appropriate; for example, if we want to model a process with time-invariant structure that is affected by seasonal errors, we might choose to make  $\mathbf{T}$ ,  $\mathbf{Z}$  time-invariant and restrict periodic behaviour to  $\mathbf{G}_t$ ,  $\mathbf{H}_t$ .

In this case, we have  $\underline{\mathbf{T}} = \mathbf{T}^s$ . However, if  $\mathbf{T}$  has eigenvalues  $\{\lambda_1, \dots, \lambda_q\}$ , then  $\mathbf{T}^s$  has eigenvalues  $\{\lambda_1^s, \dots, \lambda_q^s\}$ . Thus:



$$\begin{aligned}
\mathbf{T} \text{ stable} &\Leftrightarrow |\lambda_j| < 1 \quad \forall j \in \{1, 2, \dots, q\} \\
&\Leftrightarrow |\lambda_j|^s < 1 \quad \forall j \in \{1, 2, \dots, q\} \\
&\Leftrightarrow \mathbf{T}^s \text{ stable}
\end{aligned} \tag{5.33}$$

regardless of the structure of the measurement matrix  $\mathbf{Z}_i$  or the errors. As an illustration, consider the periodic AR(1) model

$$\begin{aligned}
\mathbf{y}_{i,t} &= \alpha_{i,t} + \mathbf{G}_i \boldsymbol{\varepsilon}_{i,t} \\
\alpha_{i,t+1} &= \begin{cases} \phi_s \alpha_{s,t} + \mathbf{H}_s \boldsymbol{\varepsilon}_{s,t} & \text{if } i = 1 \\ \phi_{i-1} \alpha_{i-1,t+1} + \mathbf{H}_{i-1} \boldsymbol{\varepsilon}_{i-1,t+1} & \text{if } i = 2, \dots, s \end{cases}
\end{aligned} \tag{5.34}$$

The transition “matrix” is the scalar quantity  $\phi_i$ , so the summary “matrix” is  $\underline{\mathbf{T}} = \phi_1 \dots \phi_s$ , the product of the autoregressive parameters over a whole period. If this parameter is time-invariant, the process is stable iff  $|\phi| < 1$ . This is exactly the same as in the (non-periodic) AR(1) model, so the parameters that define the periodic behaviour of the measurement and transition errors have no effect on the stability of the process. On the other hand, if we allow the autoregressive parameter to vary periodically, the process is stable iff  $|\phi_1 \dots \phi_s| < 1$ . This is somewhat surprising; for example, a suitable choice of  $\phi_1$  can ensure that the process is stable even if  $|\phi_i| \geq 1$  for  $i = 2, \dots, s$ .

#### 5.3.4 Detectability and Observability

In order to establish whether or not the time-invariant process is detectable, we need to consider the eigenvectors of the matrix  $\widehat{\mathbf{T}}$ . Let  $\widehat{\mathbf{u}} = [\mathbf{u}'_1 \dots \mathbf{u}'_s \mid \mathbf{v}'_1 \dots \mathbf{v}'_s]'$  be such an eigenvector and  $\lambda$  the corresponding eigenvalue. We have

$$(\widehat{T} - \lambda \mathbf{I})\widehat{\mathbf{u}} = \mathbf{0}$$

$$\Leftrightarrow \left[ \begin{array}{cccc|cccc} -\lambda \mathbf{I} & \cdots & \mathbf{0} & \mathbf{T}_1 & & & & \\ \vdots & \ddots & \vdots & \vdots & & & & \\ \mathbf{0} & \cdots & -\lambda \mathbf{I} & \mathbf{T}_{s-1} \cdots \mathbf{T}_1 & & & & \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_s \cdots \mathbf{T}_1 - \lambda \mathbf{I} & & & & \\ \hline \mathbf{0} & \cdots & \cdots & \mathbf{0} & & & & \\ \vdots & \ddots & & \vdots & & & & \\ \vdots & & \ddots & \vdots & & & & \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & & & & \\ \hline \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}_1 & & & & \\ \vdots & \ddots & \vdots & \vdots & & & & \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_{s-1} \cdots \mathbf{T}_2 \mathbf{H}_1 & & & & \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{T}_s \cdots \mathbf{T}_2 \mathbf{H}_1 & & & & \\ \hline -\lambda \mathbf{I} & \cdots & \cdots & \mathbf{0} & & & & \\ \vdots & \ddots & & \vdots & & & & \\ \vdots & & \ddots & \vdots & & & & \\ \mathbf{0} & \cdots & \cdots & -\lambda \mathbf{I} & & & & \end{array} \right] \left[ \begin{array}{c} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{s-1} \\ \mathbf{u}_s \\ \hline \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_s \end{array} \right] = \mathbf{0}$$

$$\Leftrightarrow \left[ \begin{array}{c} -\lambda \mathbf{u}_1 + \mathbf{T}_1 \mathbf{u}_s + \mathbf{H}_1 \mathbf{v}_s \\ -\lambda \mathbf{u}_2 + \mathbf{T}_2 \mathbf{T}_1 \mathbf{u}_s + \mathbf{T}_2 \mathbf{H}_1 \mathbf{v}_s \\ \vdots \\ -\lambda \mathbf{u}_s + \mathbf{T}_s \cdots \mathbf{T}_1 \mathbf{u}_s + \mathbf{T}_s \cdots \mathbf{T}_2 \mathbf{H}_1 \mathbf{v}_s \\ \hline -\lambda \mathbf{v}_1 \\ \vdots \\ -\lambda \mathbf{v}_s \end{array} \right] = \mathbf{0}$$

$$\Leftrightarrow \left. \begin{array}{l} \mathbf{u}_1 = \frac{1}{\lambda} \mathbf{T}_1 \mathbf{u}_s \\ \vdots \\ \mathbf{u}_{s-1} = \frac{1}{\lambda} \mathbf{T}_{s-1} \cdots \mathbf{T}_1 \mathbf{u}_s \\ (\underline{\mathbf{T}} \mathbf{u}_s - \lambda \mathbf{I}) = \mathbf{0} \\ \mathbf{v}_1 = \mathbf{0} \\ \vdots \\ \mathbf{v}_s = \mathbf{0} \end{array} \right\} \quad (5.35)$$

Thus,  $\mathbf{u}_s$  is an eigenvector of  $\underline{\mathbf{T}}$ . We ignore the cases  $\lambda = 0$  and  $\mathbf{u}_s = \mathbf{0}$  because, for detectability, we only need to examine eigenvectors corresponding to eigenvalues outside the unit circle. Let  $\mathbf{u}_s$  be such an eigenvector and consider

$$\begin{aligned} \widehat{\mathbf{Z}}\widehat{\mathbf{u}} &= \begin{bmatrix} \mathbf{Z}_1 & & & & \mathbf{G}_1 & & \\ & \ddots & & & & \ddots & \\ & & \mathbf{Z}_s & & & & \mathbf{G}_s \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda}\mathbf{T}_1\mathbf{u}_s \\ \vdots \\ \frac{1}{\lambda}\mathbf{T}_{s-1}\dots\mathbf{T}_1\mathbf{u}_s \\ \hline \mathbf{u}_s \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\lambda}\mathbf{Z}_1\mathbf{T}_1\mathbf{u}_s \\ \vdots \\ \frac{1}{\lambda}\mathbf{Z}_{s-1}\mathbf{T}_{s-1}\dots\mathbf{T}_1\mathbf{u}_s \\ \hline \mathbf{Z}_s\mathbf{u}_s \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \end{aligned} \quad (5.36)$$

The process is detectable iff  $\widehat{\mathbf{u}} \neq \mathbf{0}$  implies  $\widehat{\mathbf{Z}}\widehat{\mathbf{u}} \neq \mathbf{0}$ . This condition can be summarised as

$$(\widehat{\mathbf{T}}, \widehat{\mathbf{Z}}) \text{ is detectable} \Leftrightarrow \begin{cases} \forall j \in \{1, \dots, s\} \text{ s.t. } |\lambda_j(\underline{\mathbf{T}})| \geq 1 \\ \exists i \in \{1, \dots, s\} \text{ s.t. } \mathbf{Z}_i\mathbf{T}_i\dots\mathbf{T}_1\mathbf{u}_s \neq \mathbf{0} \end{cases} \quad (5.37)$$

where  $\mathbf{u}_s$  is the corresponding eigenvector.

The condition for observability is very similar, but we need to test the eigenvectors corresponding to all non-zero eigenvalues, not just the ones outside the unit circle. This can be summarised as:

$$(\widehat{\mathbf{T}}, \widehat{\mathbf{Z}}) \text{ is observable} \Leftrightarrow \begin{cases} \forall j \in \{1, \dots, s\} \text{ s.t. } \lambda_j(\underline{\mathbf{T}}) \neq 0 \\ \exists i \in \{1, \dots, s\} \text{ s.t. } \mathbf{Z}_i \mathbf{T}_i \dots \mathbf{T}_1 \mathbf{u}_s \neq 0 \end{cases} \quad (5.38)$$

where  $\mathbf{u}_s$  is the corresponding eigenvector.

### 5.3.5 Stabilisability

To determine whether or not the time-invariant process is stabilisable, we need to consider the eigenvectors of  $\widehat{\mathbf{T}}'$ . Let  $\widehat{\mathbf{w}} = [\mathbf{w}'_1 \dots \mathbf{w}'_s \mid \mathbf{z}'_1 \dots \mathbf{z}'_s]'$  be such an eigenvector and  $\lambda$  the corresponding eigenvalue. We have

$$(\widehat{\mathbf{T}}' - \lambda \mathbf{I}) \widehat{\mathbf{u}} = \mathbf{0}$$

$$\Leftrightarrow \left[ \begin{array}{cccc|cccc} -\lambda \mathbf{I} & \dots & \mathbf{0} & \mathbf{0} & & & & \\ \vdots & \ddots & \vdots & \vdots & & & & \\ \mathbf{0} & \dots & -\lambda \mathbf{I} & \mathbf{0} & & & & \\ \mathbf{T}'_1 & \dots & \mathbf{T}'_1 \dots \mathbf{T}'_{s-1} & \underline{\mathbf{T}}' - \lambda \mathbf{I} & & & & \\ \hline \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & & & & \\ \vdots & \ddots & \vdots & \vdots & & & & \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & & & & \\ \mathbf{H}'_1 & \dots & \mathbf{H}'_1 \mathbf{T}'_2 \dots \mathbf{T}'_{s-1} & \mathbf{H}'_1 \mathbf{T}'_2 \dots \mathbf{T}'_s & & & & \\ & & & & \mathbf{0} & \dots & \dots & \mathbf{0} \\ & & & & \vdots & \ddots & & \vdots \\ & & & & \vdots & & \ddots & \mathbf{0} \\ & & & & \mathbf{0} & \dots & \dots & \mathbf{0} \\ & & & & -\lambda \mathbf{I} & \dots & \dots & \mathbf{0} \\ & & & & \vdots & \ddots & & \vdots \\ & & & & \vdots & & \ddots & \vdots \\ & & & & \mathbf{0} & \dots & \dots & -\lambda \mathbf{I} \end{array} \right] \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_s \\ \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_s \end{bmatrix} = \mathbf{0}$$

$$\begin{aligned}
& \Leftrightarrow \left[ \begin{array}{c} -\lambda \mathbf{w}_1 \\ \vdots \\ -\lambda \mathbf{w}_{s-1} \\ \hline \mathbf{T}'_1 \mathbf{w}_1 + \dots + \mathbf{T}'_1 \dots \mathbf{T}'_{s-1} \mathbf{w}_{s-1} + (\mathbf{T}' - \lambda \mathbf{I}) \mathbf{w}_s \\ \hline -\lambda \mathbf{z}_1 \\ \vdots \\ -\lambda \mathbf{z}_{s-1} \\ \mathbf{H}'_1 \mathbf{w}_1 + \dots + \mathbf{H}'_1 \mathbf{T}'_2 \dots \mathbf{T}'_s \mathbf{w}_s - \lambda \mathbf{z}_s \end{array} \right] = \mathbf{0} \\
& \Leftrightarrow \left. \begin{array}{l} \mathbf{w}_1 = \mathbf{0} \\ \vdots \\ \mathbf{w}_{s-1} = \mathbf{0} \\ (\mathbf{T}' - \lambda \mathbf{I}) \mathbf{w}_s = \mathbf{0} \\ \mathbf{z}_1 = \mathbf{0} \\ \vdots \\ \mathbf{z}_{s-1} = \mathbf{0} \\ \mathbf{z}_s = \frac{1}{\lambda} \mathbf{H}'_1 \mathbf{T}'_2 \dots \mathbf{T}'_s \mathbf{w}_s \end{array} \right\} \quad (5.39)
\end{aligned}$$

This time,  $\mathbf{w}_s$  is an eigenvector of  $\mathbf{T}'$  and the eigenvector  $\hat{\mathbf{w}}$  is non-zero iff  $\mathbf{w}_s$  is non-zero. Note that the eigenvalues of  $\mathbf{T}'$  are the same as the eigenvalues of  $\mathbf{T}$ . We now consider the expression

$$\widehat{\mathbf{H}}' \widehat{\mathbf{w}}$$

$$= \begin{bmatrix} 0 & \mathbf{H}'_2 & \mathbf{H}'_2 \mathbf{T}'_3 & \cdots & \mathbf{H}'_2 \mathbf{T}'_3 \cdots \mathbf{T}'_s & \mathbf{I} \\ 0 & & \mathbf{H}'_3 & \cdots & \mathbf{H}'_3 \mathbf{T}'_4 \cdots \mathbf{T}'_s & \\ \vdots & & & \ddots & \vdots & \\ 0 & & & & \mathbf{H}'_s & \\ 0 & 0 & 0 & \cdots & 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathbf{w}_s \\ 0 \\ \vdots \\ 0 \\ \frac{1}{\lambda} \mathbf{H}'_1 \mathbf{T}'_2 \cdots \mathbf{T}'_s \mathbf{w}_s \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{H}'_2 \mathbf{T}'_3 \cdots \mathbf{T}'_s \mathbf{w}_s \\ \vdots \\ \mathbf{H}'_s \mathbf{w}_s \\ \frac{1}{\lambda} \mathbf{H}'_1 \mathbf{T}'_2 \cdots \mathbf{T}'_s \mathbf{w}_s \end{bmatrix} \quad (5.40)$$

The process is stabilisable iff  $\widehat{\mathbf{w}} \neq \mathbf{0}$  implies  $\widehat{\mathbf{H}}' \widehat{\mathbf{w}} \neq \mathbf{0}$ . We can hence summarise the condition as:

$$(\widehat{\mathbf{T}}, \widehat{\mathbf{H}}) \text{ is stabilisable} \Leftrightarrow \begin{cases} \forall j \in \{1, \dots, s\} \text{ s.t. } |\lambda_j(\underline{\mathbf{T}}')| \geq 1 \\ \exists i \in \{1, \dots, s\} \text{ s.t. } \mathbf{H}'_i \mathbf{T}'_{i+1} \cdots \mathbf{T}'_s \mathbf{w}_s \neq \mathbf{0} \end{cases} \quad (5.41)$$

where  $\mathbf{w}_s$  is the corresponding eigenvector of  $\underline{\mathbf{T}}'$ . Similarly, the condition for controllability is:

$$(\widehat{\mathbf{T}}, \widehat{\mathbf{H}}) \text{ is controllable} \Leftrightarrow \begin{cases} \forall j \in \{1, \dots, s\} \text{ s.t. } \lambda_j(\underline{\mathbf{T}}') \neq 0 \\ \exists i \in \{1, \dots, s\} \text{ s.t. } \mathbf{H}'_i \mathbf{T}'_{i+1} \cdots \mathbf{T}'_s \mathbf{w}_s \neq \mathbf{0} \end{cases} \quad (5.42)$$

where  $\mathbf{w}_s$  is the corresponding eigenvector of  $\underline{\mathbf{T}}'$ .

## 5.4 Conclusions

Steady-state convergence is an important computational advantage of the Kalman filter. In most practical examples of time-invariant models, the matrix  $\mathbf{P}_t$  tends to converge to a constant value very quickly, rendering part of the filter recursions redundant.

The method described in this chapter makes it possible to benefit from the computational efficiency of the UR and the steady-state filter simultaneously. This can lead to considerable savings, both in the number of operations involved and in the amount of stored data required for the backward pass of the KFS. The improvement is particularly noticeable in long multivariate time series, such as the temperature data in Chapter 6.

The technique of incorporating the error term into the state vector serves to demonstrate the equivalence between the control theory state-space model and the general time series SSF. We have shown that this equivalence provides rigorous justification for the use of the standard convergence conditions.

The novel concept of periodic convergence allows us to extend the computational benefit of steady-state convergence to a wide class of seasonal models. The set of conditions established in the last part of this chapter provide a framework for the treatment of these models.

## Chapter 6

# Estimation of Missing Temperature Data

### 6.1 Introduction

Serially complete and reliable temperature records are essential for the detection of global climate change and are also required for the development of climate-dependent models for soil erosion, crop development and other such processes (DeGaetano *et al.*, 1995). Furthermore, without a large amount of high-quality weather data, management of weather risk and pricing of weather derivatives would be unfeasible (Dunis & Karalis, 2003). However, the gaps that are often encountered in long temperature series are a serious hindrance to these endeavours. In addition, climatic models can be extremely sensitive to outliers and errant values, which can arise in the data from a variety of sources.

In the past, missing values were replaced arbitrarily or with crude estimation techniques, which obviously affected the accuracy of the final models and impaired the comparison of results obtained with different modelling approaches. This has led researchers to develop a large number of methods which produce far more realistic daily temperature estimates. The estimated data can then be used to fill gaps in the records and also to identify outliers by drawing attention to cases where there are large discrepancies between observed and estimated temperature readings.



In the first part of this chapter, we begin by reviewing some of the most important existing approaches for estimation of missing temperature observations (Kemp *et al.*, 1983; De Gaetano *et al.*, 1995). These tend to be non-parametric methods, and do not make any attempt to model the temperature series directly. We examine the relative accuracy of different estimation techniques and propose a simple improvement which can lead more accurate results, without a significant increase in computational cost. Additionally, we consider the use of the Nearest-Neighbours approach (Edelsbrunner, 1987; Kleinberg, 1997; Indyk & Motwani, 1999) and introduce novel approximate methods for nearest-neighbour search, which go some way towards alleviating the “curse of dimensionality”.

In the second part of the chapter, we consider a model-based approach to missing temperature estimation. Seasonal models are used to remove the long-term cyclical (climate) patterns from the series, thus enabling us to focus on the more volatile weather effects. These weather series tend to show evidence of long-memory behaviour (Caballero *et al.*, 2001). We consider heuristic and likelihood-based methods for filtering out long memory (Hosking, 1981; Beran, 1989 & 1994; Taqqu *et al.*, 1995). The resulting series can be adequately modelled using the autoregressive models discussed in the technical introduction. We focus on fitting multivariate models to several series simultaneously, which allows us to improve estimation accuracy by exploiting the high degree of correlation between temperature series at locations that are near each other.

All models are then applied to a long multivariate time series, consisting of daily maximum temperatures at weather stations in the state of Texas, from 1950 to 2001. Numerical results are given in the Appendix.

### 6.1.1 Definitions

The problem of estimating missing daily temperatures can be formulated as follows:

We observe  $p$  contemporaneous univariate time series. Let  $\mathbf{Y}_i$  denote the stacked observations of the  $i^{\text{th}}$  series:

$$\{\mathbf{Y}\}_i = (y_{i,1}, y_{i,2}, \dots, y_{i,n})' \quad i = 1, \dots, p \quad (6.1)$$

a  $n \times 1$  vector. The  $\{\mathbf{Y}\}_i$  are the columns of the  $n \times p$  matrix  $\mathbf{Y}$ , defined as

$$\mathbf{Y} = \begin{bmatrix} \{\mathbf{Y}\}_1 & \{\mathbf{Y}\}_2 & \cdots & \{\mathbf{Y}\}_p \end{bmatrix} \quad (6.2)$$

Using the previous definitions, we can also express  $\mathbf{Y}$  as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}'_1 \\ \mathbf{y}'_2 \\ \vdots \\ \mathbf{y}'_n \end{bmatrix} \quad (6.3)$$

where  $\mathbf{y}_t$  is the  $p \times 1$  vector of observations at time  $t$ . The stacked vector of observations  $\mathbf{y}$  is equal to  $\text{vec}(\mathbf{Y})$ .

Finally, we need to define the  $(n-1) \times 1$  vector

$$\{\mathbf{Y}\}_{i \setminus t} = (y_{i,1}, \dots, y_{i,t-1}, y_{i,t+1}, \dots, y_{i,n})' \quad (6.4)$$

the  $(p-1) \times 1$  vector

$$\mathbf{y}_{t \setminus i} = (y_{1,t}, \dots, y_{i-1,t}, y_{i+1,t}, \dots, y_{p,t})' \quad (6.5)$$

which are the same as  $\{\mathbf{Y}\}_i$  and  $\mathbf{y}_t$ , respectively, but with observation  $y_{i,t}$  omitted, and the  $(np-1) \times 1$  stack vector

$$\mathbf{y}_{\setminus i,t} = (\mathbf{y}'_1, \dots, \mathbf{y}'_{t-1}, \mathbf{y}'_{t \setminus i}, \mathbf{y}'_{t+1}, \dots, \mathbf{y}'_n)' \quad (6.6)$$

which is the same as  $\mathbf{y}$ , but with observation  $y_{i,t}$  omitted. We will consider the problem where a single observation  $y_{i,t}$  is missing and needs to be estimated.

## 6.2 Ad-hoc Approach

### 6.2.1 Existing Methods

Kemp *et al.* (1983) make an early attempt to classify and compare different methods for estimating maximum and minimum daily temperatures. They examined several methods, which they divide into three broad categories: 1) within-station, 2) between-station, and 3) regression-based.

#### Within-Station Methods

Within-station methods involve estimating a missing observation by only taking into account temperatures recorded on previous and subsequent days at the same weather station. For example, the maximum temperature on 15 May could be estimated by the average of the maximum temperatures on 14 May and 16 May. Similarly, it is possible to calculate averages by considering more than one day on either side of the missing observation, or by assigning different weights to the days. Although these methods can produce satisfactory results when calculating monthly or longer period averages, they tend to result in large estimation errors for daily temperatures and are generally deemed inadequate.

#### Between-Station Methods: Temperature Departures

Between-station methods are multivariate approaches to estimating missing observations. Broadly speaking, they take advantage of the correlation between the components of  $\mathbf{y}_t$  to improve the estimate of  $y_{i,t}$ .

For example, Kemp *et al.* (1983) describe a procedure which is based on the assumption that daily temperatures in neighbouring stations differ, on average, by an amount equal to the difference between their corresponding average monthly temperatures. DeGaetano *et al.* (1995) use a more accurate method which involves calculating *standard departures* for each observation.

The standard departure  $z_{i,t}^\dagger$  for station  $i$  on day  $t$  is defined as:

$$z_{i,t}^\dagger \equiv \frac{y_{i,t} - \bar{y}_{i,t}}{s_{i,t}} \quad (6.7)$$

where  $y_{i,t}$  is the daily maximum (or minimum) temperature at station  $i$  on that day and  $\bar{y}_{i,t}$ ,  $s_{i,t}$  are estimates of the mean daily maximum (or minimum) for that day and its standard deviation, respectively. In practice, estimates of the monthly means and standard deviations are used to calculate the daily means and standard deviations. For days in the second and third weeks of each month  $\bar{y}_{i,t}$ ,  $s_{i,t}$  are taken to be equal to the corresponding monthly estimates, while for days in the first (or last) week they are obtained by averaging the estimates for the current and preceding (or following) month. The standard departure  $\hat{z}_{i,t}^\dagger$  for the station with missing data is then estimated as

$$\hat{z}_{i,t}^\dagger \equiv \frac{1}{p-1} \sum_{\substack{j=1 \\ j \neq i}}^p z_{j,t}^\dagger \quad (6.8)$$

where the sum is over all neighbouring stations,  $p-1$  stations in total. The estimate of the missing data  $\hat{y}_{i,t}^\dagger$  is:

$$\hat{y}_{i,t}^\dagger \equiv s_{i,t} \hat{z}_{i,t}^\dagger + \bar{y}_{i,t} \quad (6.9)$$

where  $\bar{y}_{i,t}$ ,  $s_{i,t}$  are the corresponding daily mean and standard deviation at the target station. Clearly, the choice of stations to include in the

model is of great importance to the overall accuracy of the estimate. DeGaetano *et al.* (1995) start by looking for any weather stations within  $0.1^\circ$  of latitude radius of the target station, then increase the search radius until at least three stations are found or the radius exceeds  $1^\circ$ . They demonstrate that this is a significantly better approach than older methods where stations are selected based on political boundaries.

### Regression-based Methods: Least Absolute Deviations

With regression-based methods, missing observations are estimated by fitting a regression model which uses the temperatures at neighbouring stations as predictors. Kemp *et al.* (1983) and Eischeid *et al.* (1995) consider a number of more robust regression criteria in addition to ordinary least squares and conclude that least absolute deviation (LAD) regression produces the best results. LAD regression (also known as L1 regression) is a robust version of ordinary least squares and consists of choosing the parameter estimates which minimize the sum of the absolute deviations (rather than the squared deviations) of  $y$  from the predicted values. The model equation is

$$y_{i,t} = \mathbf{y}_{t \setminus i} \boldsymbol{\beta} + \eta_t, \quad t = 1, \dots, n \quad (6.10)$$

where  $\boldsymbol{\beta}$  is a  $(p-1) \times 1$  vector of parameters. The errors  $\eta_t$  are assumed to be independent and have zero mean and constant variance. The parameters  $\boldsymbol{\beta}$  are estimated according to the LAD criterion:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \sum_{\substack{s=1 \\ s \neq t}}^n \left| y_{i,s} - \mathbf{y}_{s \setminus i} \boldsymbol{\beta} \right| \quad (6.11)$$

The resulting estimate for  $y_{i,t}$  is then

$$\hat{y}_{i,t} = \mathbf{y}_{t \setminus i} \hat{\boldsymbol{\beta}} \quad (6.12)$$

The LAD method is implemented using the algorithms of Barrodale & Roberts (1973). The spatial structure of the series can change drastically over the course of the year. For example, depending on the direction of warm/cold fronts, a particular neighbouring station may be the best predictor for the target station in December but not in July. To account for this, 12 different sets of regression coefficients  $\hat{\beta}$  were calculated — one for each month.

### 6.2.2 The Harmonic Model

#### Harmonic Component

Since the temperature series exhibit strong periodic behaviour, it is reasonable to model the long-term cycles with a trigonometric component. We first subtract the average temperature, to make the mean of the series equal to zero, then consider the simple trigonometric model:

$$\begin{aligned} u_t &= \mu_t + \zeta_t \\ \mu_t &= R \cos(\omega t + \phi), \quad t = 0, \dots, N-1 \end{aligned} \quad (6.13)$$

Periodograms of daily temperature series have a very strong peak at a frequency of  $\omega = 2\pi/T \approx 0.0172$ , which is an indication of cyclical behaviour in the data (Bloomfield, 1976). In this case, the highest peak corresponds to a period of  $T \approx 365.2422$  observations — the number of days in the mean tropical year, as expected (Figure A.1). We will now proceed to show how the model parameters can be estimated using the equivalent form

$$\mu_t = A \cos \omega t + B \sin \omega t \quad (6.14)$$

The coefficients  $A$  and  $B$  can be estimated by minimizing the sum of squares:

$$S(A, B) = \sum_{t=0}^{N-1} (u_t - A \cos \omega t - B \sin \omega t)^2 \quad (6.15)$$

The partial derivatives with respect to the two unknowns are

$$\begin{aligned} \frac{\partial S}{\partial A} &= -2 \sum_{t=0}^{N-1} \cos \omega t (u_t - A \cos \omega t - B \sin \omega t) \\ \frac{\partial S}{\partial B} &= -2 \sum_{t=0}^{N-1} \sin \omega t (u_t - A \cos \omega t - B \sin \omega t) \end{aligned} \quad (6.16)$$

Setting these equal to zero produces the solutions:

$$\begin{aligned} \hat{A} &= \frac{1}{\Delta} \left[ \left( \sum_{t=0}^{N-1} u_t \cos \omega t \right) \left( \sum_{t=0}^{N-1} \sin^2 \omega t \right) \right. \\ &\quad \left. - \left( \sum_{t=0}^{N-1} u_t \sin \omega t \right) \left( \sum_{t=0}^{N-1} \cos \omega t \sin \omega t \right) \right] \\ \hat{B} &= \frac{1}{\Delta} \left[ \left( \sum_{t=0}^{N-1} u_t \sin \omega t \right) \left( \sum_{t=0}^{N-1} \cos^2 \omega t \right) \right. \\ &\quad \left. - \left( \sum_{t=0}^{N-1} u_t \cos \omega t \right) \left( \sum_{t=0}^{N-1} \cos \omega t \sin \omega t \right) \right] \end{aligned} \quad (6.17)$$

where

$$\Delta = \left( \sum_{t=0}^{N-1} \sin^2 \omega t \right) \left( \sum_{t=0}^{N-1} \cos^2 \omega t \right) - \left( \sum_{t=0}^{N-1} \cos \omega t \sin \omega t \right)^2$$

To evaluate the trigonometric terms in  $\Delta$ , consider

$$\begin{aligned}
& \sum_{t=0}^{N-1} (\cos(\omega t) + i \sin(\omega t)) \\
&= \sum_{t=0}^{N-1} \exp(i\omega t) \\
&= \frac{\exp(iN\omega) - 1}{\exp(i\omega) - 1} \\
&= \exp\left(\frac{i(N-1)\omega}{2}\right) \frac{\exp(iN\omega/2) - \exp(-iN\omega/2)}{\exp(i\omega/2) - \exp(-i\omega/2)} \\
&= \left[ \cos\left(\frac{(N-1)\omega}{2}\right) + i \sin\left(\frac{(N-1)\omega}{2}\right) \right] \frac{2i \sin(N\omega/2)}{2i \sin(\omega/2)} \quad (6.18)
\end{aligned}$$

where  $i = \sqrt{-1}$ . Comparing the real and imaginary parts yields

$$\begin{aligned}
\sum_{t=0}^{N-1} \cos(\omega t) &= N \cos\left(\frac{(N-1)\omega}{2}\right) D_N(\omega) \\
\sum_{t=0}^{N-1} \sin(\omega t) &= N \sin\left(\frac{(N-1)\omega}{2}\right) D_N(\omega) \quad (6.19)
\end{aligned}$$

where

$$D_N(\omega) = \frac{\sin(N\omega/2)}{N \sin(\omega/2)} \quad (6.20)$$

Now the terms can be computed:



$$\begin{aligned}\sum_{t=0}^{N-1} \cos^2 \omega t &= \frac{1}{2} \sum_{t=0}^{N-1} [\cos(2\omega t) + 1] \\ &= \frac{N}{2} [1 + D_N(2\omega) \cos(N-1)\omega]\end{aligned}$$

$$\begin{aligned}\sum_{t=0}^{N-1} \cos \omega t \sin \omega t &= \frac{1}{2} \sum_{t=0}^{N-1} \sin(2\omega t) \\ &= \frac{N}{2} D_N(2\omega) \sin(N-1)\omega\end{aligned}$$

$$\begin{aligned}\sum_{t=0}^{N-1} \sin^2 \omega t &= \frac{1}{2} \sum_{t=0}^{N-1} [1 - \cos(2\omega t)] \\ &= \frac{N}{2} [1 - D_N(2\omega) \cos(N-1)\omega]\end{aligned}\quad (6.21)$$

If the frequency  $\omega$  is known,  $D_N(\omega)$  can be evaluated exactly. In the general case, since  $|ND_N(\omega)| \leq 1/\sin(\omega/2)$ , the terms involving  $D_N$  are all small compared to  $N/2$ , provided that  $N$  is large and  $\omega$  is not too close to zero. We can thus omit these terms to obtain a simpler set of equations (Bloomfield, 1976):

$$\begin{aligned}\tilde{A} &= \frac{2}{N} \sum_{t=0}^{N-1} u_t \cos \omega t \\ \tilde{B} &= \frac{2}{N} \sum_{t=0}^{N-1} u_t \sin \omega t\end{aligned}\quad (6.22)$$

## Multiple Periodicities

Constructing more complex periodic models is straightforward if the simple model is expanded to:

$$\begin{aligned}
u_t &= \mu_t + \zeta_t \\
\mu_t &= \sum_{j=1}^h R_j \cos(\omega_j t + \phi_j), \quad t = 0, \dots, N-1
\end{aligned} \tag{6.23}$$

for  $h$  frequencies. As in the simple trigonometric case, this can be written as:

$$\mu_t = \sum_{j=1}^h (A_j \cos \omega_j t + B_j \sin \omega_j t) \tag{6.24}$$

There is a similar set of approximations that can be used in this case — consider the sum of squares:

$$S = \sum_{t=0}^{N-1} \left\{ u_t - \sum_{j=1}^h (A_j \cos \omega_j t + B_j \sin \omega_j t) \right\}^2 \tag{6.25}$$

This yields the following partial derivatives:

$$\begin{aligned}
\frac{\partial S}{\partial A_j} &= -2 \sum_{t=0}^{N-1} \cos \omega_j t \left( u_t - \sum_{i=1}^h (A_i \cos \omega_i t + B_i \sin \omega_i t) \right) \\
\frac{\partial S}{\partial B_j} &= -2 \sum_{t=0}^{N-1} \sin \omega_j t \left( u_t - \sum_{i=1}^h (A_i \cos \omega_i t + B_i \sin \omega_i t) \right)
\end{aligned} \tag{6.26}$$

for  $j = 1, 2, \dots, h$ . Terms of the form  $\sum \cos^2 \omega_j t$ ,  $\sum \cos \omega_j t \sin \omega_j t$  and  $\sum \sin^2 \omega_j t$  can be computed as before. Consider the ‘cross’ terms, i.e. those involving summations of trigonometric functions of two different frequencies. They can be computed using standard trigonometric formulae and the previous results:

$$\begin{aligned}
& \sum_{t=0}^{N-1} \cos \omega_i t \cos \omega_j t \\
&= \frac{1}{2} \sum_{t=0}^{N-1} (\cos \Omega_{ij} t + \cos \Omega_{(ij)} t) \\
&= \frac{1}{2} \left[ \cos \left( \frac{(N-1)\Omega_{ij}}{2} \right) D_N(\Omega_{ij}) + \cos \left( \frac{(N-1)\Omega_{(ij)}}{2} \right) D_N(\Omega_{(ij)}) \right]
\end{aligned} \tag{6.27}$$

$$\begin{aligned}
& \sum_{t=0}^{N-1} \sin \omega_i t \sin \omega_j t \\
&= \frac{1}{2} \sum_{t=0}^{N-1} (\cos \Omega_{(ij)} t - \cos \Omega_{ij} t) \\
&= \frac{1}{2} \left[ \cos \left( \frac{(N-1)\Omega_{(ij)}}{2} \right) D_N(\Omega_{(ij)}) - \cos \left( \frac{(N-1)\Omega_{ij}}{2} \right) D_N(\Omega_{ij}) \right]
\end{aligned} \tag{6.28}$$

$$\begin{aligned}
& \sum_{t=0}^{N-1} \cos \omega_i t \sin \omega_j t \\
&= \frac{1}{2} \sum_{t=0}^{N-1} (\sin \Omega_{ij} t - \sin \Omega_{(ij)} t) \\
&= \frac{1}{2} \left[ \sin \left( \frac{(N-1)\Omega_{ij}}{2} \right) D_N(\Omega_{ij}) - \sin \left( \frac{(N-1)\Omega_{(ij)}}{2} \right) D_N(\Omega_{(ij)}) \right]
\end{aligned} \tag{6.29}$$

where  $\Omega_{ij} = \omega_i + \omega_j$ ,  $\Omega_{(ij)} = \omega_i - \omega_j$  and  $i \neq j$ . These terms all involve  $D_N$  hence, under similar assumptions as before, they are small compared with  $N/2$  and can be ignored. Thus, the partial derivatives are reduced to

$$\begin{aligned}\frac{\partial S}{\partial A_j} &= -2 \sum_{t=0}^{N-1} \cos \omega_j t (u_t - A_j \cos \omega_j t - B_j \sin \omega_j t) \\ \frac{\partial S}{\partial B_j} &= -2 \sum_{t=0}^{N-1} \sin \omega_j t (u_t - A_j \cos \omega_j t - B_j \sin \omega_j t)\end{aligned}\quad (6.30)$$

which yield the approximate solutions:

$$\begin{aligned}\tilde{A}_j &= \frac{2}{N} \sum_{t=0}^{N-1} u_t \cos \omega_j t \\ \tilde{B}_j &= \frac{2}{N} \sum_{t=0}^{N-1} u_t \sin \omega_j t\end{aligned}\quad (6.31)$$

In this model,  $\omega_1 = 0.0172$  is the fundamental frequency,  $\omega_2 = 2\omega_1$  the first harmonic,  $\omega_3 = 3\omega_1$  the second harmonic, etc. depending on the number of harmonics included in the model,  $h$ , which is chosen arbitrarily. In practice, a total of two or three frequencies are required in order to remove the seasonality from the monthly means (Figure A.2).

The resulting model is thus

$$\begin{aligned}y_{i,t} &= \alpha_i + \mu_{i,t} + \zeta_{i,t} \\ \mu_{i,t} &= \sum_{j=1}^h R_{i,j} \cos(\omega_{i,j}t + \phi_{i,j}), \quad t = 1, \dots, n\end{aligned}\quad (6.32)$$

where the constant term  $\alpha_i$  is the mean of the series  $\{\mathbf{Y}\}_i$ .

## Linear Trend

In addition to the trigonometric component, it is sensible to include a linear trend in order to model possible long-term shifts in the underlying mean. The resulting model, assuming there is no interaction between the trend and harmonic components, can be written as:

$$y_{i,t} = \alpha_i + \beta_i t + \sum_{j=1}^h R_{i,j} \sin(\omega_{i,j} t - \phi_{i,j}) + \zeta_{i,t} \quad (6.33)$$

for  $t = 1, \dots, n$ , where  $\zeta_{i,t}$  is a process with zero mean and variance  $\sigma_{i,t}^2$ . Again, the model can be fitted using the equivalent form:

$$y_{i,t} = \alpha_i + \beta_i t + \sum_{j=1}^h \{c_{i,j} \cos(\omega_{i,j} t) + d_{i,j} \sin(\omega_{i,j} t)\} + \zeta_{i,t} \quad (6.34)$$

for  $t = 1, \dots, n$ . Parameters  $\alpha_i, \beta_i, c_{i,1}, \dots, c_{i,h}$  and  $d_{i,1}, \dots, d_{i,h}$  (a total of  $l = 2h + 2$  parameters) need to be estimated. The formulae in the previous section cannot be applied, because of the introduction of the slope term, so we need to use least squares regression. By defining the  $l \times 1$  parameter vector  $\gamma_i = (\alpha_i, \beta_i, c_{i,1}, d_{i,1}, \dots, c_{i,h}, d_{i,h})'$ , the model equation can be written in matrix notation as:

$$\{Y\}_i = X\gamma_i + \zeta_i \quad (6.35)$$

where  $X$  is the  $n \times l$  matrix:

$$X = \begin{pmatrix} 1 & 1 & \cos \frac{2\pi}{T} & \sin \frac{2\pi}{T} & \cdots & \cos \frac{2h\pi}{T} & \sin \frac{2h\pi}{T} \\ 1 & 2 & \cos \frac{2\pi}{T} 2 & \sin \frac{2\pi}{T} 2 & \cdots & \cos \frac{2h\pi}{T} 2 & \sin \frac{2h\pi}{T} 2 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & n & \cos \frac{2\pi}{T} n & \sin \frac{2\pi}{T} n & \cdots & \cos \frac{2h\pi}{T} n & \sin \frac{2h\pi}{T} n \end{pmatrix} \quad (6.36)$$

Thus, the parameter estimates are given by  $\hat{\gamma}_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{\mathbf{Y}\}_i$ . Referring back to equation (6.34), the original parameters  $R_{i,1}, \dots, R_{i,h}$  and  $\phi_{i,1}, \dots, \phi_{i,h}$  can now be estimated using the identities

$$\begin{aligned} R_{i,j}^2 &= c_{i,j}^2 + d_{i,j}^2 \\ \tan \phi_{i,j} &= -\frac{c_{i,j}}{d_{i,j}} \end{aligned} \quad (6.37)$$

for  $j = 1, \dots, h$ . We obtain a unique estimate for  $\phi_{i,j}$  by introducing the condition  $\phi_{i,j} \in [0, \pi)$ .

### Trigonometric Model with Interactions

It is also possible to allow for interaction between the trend and seasonal components. A suitable formulation of the model is

$$y_{i,t} = \alpha_i + \beta_i t + \sum_{j=1}^h a_{i,j} \sin(\omega_{i,j}t - \phi_{i,j}) + \sum_{j=1}^h b_{i,j} t \sin(\omega_{i,j}t - \phi_{i,j}) + \zeta_{i,t} \quad (6.38)$$

for  $t=1, \dots, n$ . In this model, the dependent variable cannot be written as a matrix product of the independent variables and the parameters, so, unlike in the previous case, the parameters cannot be estimated with least squares regression. However, we can sidestep this issue by replacing the phase parameters  $\phi_{i,1}, \dots, \phi_{i,h}$  with their estimates  $\hat{\phi}_{i,1}, \dots, \hat{\phi}_{i,h}$  from the previous model and then treating them as constants. The matrix form of the regression equation is now

$$\{\mathbf{Y}\}_i = \hat{\mathbf{X}}_i^* \boldsymbol{\gamma}_i^* + \boldsymbol{\zeta}_i \quad (6.39)$$

where  $\boldsymbol{\gamma}_i^* = (\alpha_i, \beta_i, a_{i,1}, \dots, a_{i,h}, b_{i,1}, \dots, b_{i,h})'$ , a  $l \times 1$  vector, and  $\hat{\mathbf{X}}_i^*$  is the  $n \times l$  matrix

$$\hat{\mathbf{X}}_i^* = \begin{pmatrix} 1 & 1 & \sin(\frac{2\pi}{T} - \hat{\phi}_{i,1}) & \cdots & \sin(\frac{2\pi}{T} - \hat{\phi}_{i,h}) \\ 1 & 2 & \sin(\frac{2\pi}{T} 2 - \hat{\phi}_{i,1}) & \cdots & \sin(\frac{2\pi}{T} 2 - \hat{\phi}_{i,h}) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & n & \sin(\frac{2\pi}{T} n - \hat{\phi}_{i,1}) & \cdots & \sin(\frac{2\pi}{T} n - \hat{\phi}_{i,h}) \end{pmatrix} \begin{pmatrix} \sin(\frac{2\pi}{T} - \hat{\phi}_{i,1}) & \cdots & \sin(\frac{2\pi}{T} - \hat{\phi}_{i,h}) \\ 2 \sin(\frac{2\pi}{T} - \hat{\phi}_{i,1}) & \cdots & 2 \sin(\frac{2\pi}{T} - \hat{\phi}_{i,h}) \\ \vdots & & \vdots \\ n \sin(\frac{2\pi}{T} - \hat{\phi}_{i,1}) & \cdots & n \sin(\frac{2\pi}{T} - \hat{\phi}_{i,h}) \end{pmatrix} \quad (6.40)$$

The parameter estimates are given by  $\hat{\gamma}_i^* = [(\hat{\mathbf{X}}_i^*)' \hat{\mathbf{X}}_i^*]^{-1} (\hat{\mathbf{X}}_i^*)' \{\mathbf{Y}\}_i$ , which yields the estimated values  $\{\hat{\mathbf{Y}}\}_i^* = \hat{\mathbf{X}}_i^* \hat{\gamma}_i^*$  and the residuals  $\zeta_i^* = \{\mathbf{Y}\}_i - \{\hat{\mathbf{Y}}\}_i^*$ .

### Deseasonalising the Variance

The residuals  $\zeta_{i,t}^*$  still exhibit strong seasonal behaviour in the second moment, since winter temperatures usually have higher variance than summer temperatures. This can be deduced from the fact that temperature series tend to have peaks in the periodogram of  $(\zeta_{i,t}^*)^2$  (Figure A.3). Thus, it is necessary to fit another harmonic model, this time to the variance terms  $\sigma_{i,t}^2$ . Let the model be

$$\sigma_{i,t}^2 = \alpha_{\sigma_i} + \sum_{j=1}^{h_\sigma} \{c_{\sigma_{i,j}} \cos(\omega_{i,j}t) + d_{\sigma_{i,j}} \sin(\omega_{i,j}t)\} + \eta_{i,t}, \quad t = 1, \dots, n \quad (6.41)$$

Since the  $\zeta_{i,t}^*$  have zero mean and variance  $\sigma_{i,t}^2$ , this model can be fitted to the squares of the residuals  $(\zeta_{i,t}^*)^2$ . The parameters are estimated in a similar way to model (6.34) and, once estimates  $\hat{\sigma}_{i,t}^2$  are obtained, the  $\zeta_{i,t}^*$  are deseasonalised by dividing them by their estimated standard deviation. Denote the residuals from this model as  $z_{i,t}$ , where

$$z_{i,t} = \frac{\zeta_{i,t}^*}{\hat{\sigma}_{i,t}}, \quad t = 1, \dots, n \quad (6.42)$$

Once again, only a small number of frequencies are required, as can be deduced from the periodogram of  $z_{i,t}$  for different values of  $h_\sigma$  (Figure A.4).

### The Proposed Method – Harmonic Departures

After fitting both parts of the harmonic model to the data, temperature estimates can be obtained in a similar way to the departure method. We introduce the *harmonic departure*  $z_{i,t}$  of observation  $y_{i,t}$ ; this is the residual from the full trigonometric model at time  $t$ , location  $i$ . The harmonic departure for the station with missing data is then estimated as

$$\hat{z}_{i,t} = \frac{1}{p-1} \sum_{\substack{j=1 \\ j \neq i}}^p z_{j,t} \quad (6.43)$$

The estimate of the missing data  $\hat{y}_{i,t}$  is thus

$$\hat{y}_{i,t} = \hat{\mu}_{i,t} + \hat{\sigma}_{i,t} \hat{z}_{i,t} \quad (6.44)$$

where

$$\hat{\mu}_{i,t} = \hat{\alpha}_i + \hat{\beta}_i t + \sum_{j=1}^h \hat{a}_{i,j} \sin(\omega_{i,j} t - \hat{\phi}_{i,j}) + \sum_{j=1}^h \hat{b}_{i,j} t \sin(\omega_{i,j} t - \hat{\phi}_{i,j}) \quad (6.45)$$

In direct correspondence to the departures method,  $\hat{\mu}_{i,t}$  and  $\hat{\sigma}_{i,t}$  are the



harmonic model estimates for the mean and standard deviation on day  $t$  at location  $i$ .

### 6.2.3 $k$ -Nearest Neighbour Methods

#### Introduction

A different approach to estimating daily temperatures is through  $k$ -Nearest Neighbour (NN) methods. As a nonparametric method, NN has an important advantage over Temperature Departures or LAD regression in that it relies on a far less rigid set of assumptions. In all of these methods, filling in the missing data in temperature records is essentially treated as a problem in function approximation, that is, we assume there exists a function  $f$  such that

$$y_{i,t} = f(\mathbf{y}_{\setminus i,t}) \quad (6.46)$$

and estimate this function.

LAD assumes that  $f(\mathbf{y}_{\setminus i,t})$  is linear in its inputs (or is well-approximated by such a function), while the Departures methods places a similar condition on the deseasonalised inputs and response.  $k$ -Nearest Neighbours only assumes that  $f(\mathbf{y}_{\setminus i,t})$  is well-approximated by a locally constant function.

#### Basic $k$ -Nearest Neighbours

In the simplest form of  $k$ -NN, we define *days*  $\mathbf{v}_t$  as vectors of the form  $\mathbf{v}_t = \mathbf{y}_{t \setminus i}$  and use a distance metric to find the  $k$  most similar days to the one of interest. The set of these days is the *neighbourhood*  $N_k(\mathbf{v}_t)$  of day  $\mathbf{v}_t$ , its set of nearest neighbours:

$$N_k(\mathbf{v}_t) \equiv \{\mathbf{v}_s : d(\mathbf{v}_t, \mathbf{v}_s) \leq D(\mathbf{v}_t)_{(k)}\} \quad (6.47)$$

where  $d(\mathbf{v}_t, \mathbf{v}_s)$  is the distance between  $\mathbf{v}_t$  and  $\mathbf{v}_s$ . The most commonly used metric is Euclidean distance:

$$d(\mathbf{v}_t, \mathbf{v}_s) = \left( \sum_{\substack{j=1 \\ j \neq i}}^p (y_{j,t} - y_{j,s})^2 \right)^{1/2} \quad (6.48)$$

though there are situations where other measures, e.g. max distance, Mahalanobis distance or Manhattan distance, might be more suitable.

$D(\mathbf{v}_t)$  is the set of distances between  $\mathbf{v}_t$  and all other days in the dataset:

$$D(\mathbf{v}_t) = \{d(\mathbf{v}_t, \mathbf{v}_s), s = 1, \dots, n, s \neq t\} \quad (6.49)$$

and  $D(\mathbf{v}_t)_{(k)}$  is the  $k^{th}$  smallest distance in  $D(\mathbf{v}_t)$ . The observation  $y_{i,t}$  can then be estimated as the average of the  $y_{i,s}$  corresponding to this neighbourhood:

$$\hat{y}_{i,t} = \frac{1}{k} \sum_{s: \mathbf{v}_s \in N_k(\mathbf{v}_t)} y_{i,s} \quad (6.50)$$

## Computational Considerations

Using Euclidean distance, the problem of obtaining the neighbourhood of  $\mathbf{v}_t$  is equivalent to the problem of, given a set of points in  $\mathbb{R}^{p-1}$ -space, finding the  $k$  nearest ones to a particular point; this is a standard problem in analytic geometry and is of major importance to several different applications, such as data compression, pattern recognition (e.g. handwriting classification), data mining and machine learning.

The most straightforward approach is to calculate and sort all the distances in  $D(\mathbf{v}_t)$ . This can be very time-consuming and tends to make the method very slow and hence unsuitable for large datasets. There are a number of different ways for dealing with this problem, depending

on the size of  $p - 1$ . If the number of dimensions is not too large (generally speaking, smaller than  $\log n$ ) there are methods for obtaining the neighbourhood which can drastically reduce the number of operations and hence the time required (Edelsbrunner, 1987).

In higher dimensions, the current solutions are far less satisfactory due to the “curse of dimensionality”. One of the more successful recent approaches (Kleinberg, 1997) makes use of a simple geometric fact: if we project points onto a line then, generally speaking, points with projections that are close to each other are likelier to be near in  $\mathbb{R}^{p-1}$ -space than points whose projections are far. Because of this, it is possible to reduce the search time by projecting the points onto a series of random lines through the origin and giving priority in the search to points whose projections are near the projection of the target point.

### **$\epsilon$ -Approximate Nearest Neighbours**

It is also possible to simplify the problem by relaxing the condition on points entering the neighbourhood of  $\mathbf{v}_t$ . Because of the similarity between many days in the dataset, replacing, say, the  $2^{nd}$  most similar day with the  $10^{th}$  would not seriously affect the estimate. Of course, this is only the case in relatively low dimensions; as the number of predictors increases, the training sample populates the input space sparsely, so the nearest neighbours are not necessarily very near at all.

A well-established method is to look for  $\epsilon$ -Approximate Nearest Neighbours ( $\epsilon$ -ANN) of the input. We say that  $\mathbf{v}_s$  is an  $\epsilon$ -ANN of  $\mathbf{v}_t$  if  $d(\mathbf{v}_t, \mathbf{v}_s) \leq (1 + \epsilon)D(\mathbf{v}_t)_{(k)}$ . Indyk & Motwani (1999) propose an  $\epsilon$ -ANN algorithm, also based on random projections, which avoids the common problem of exponential dependence on  $p - 1$ .

### **The Seasonal Window Method**

We introduce a different approach to simplifying the search by exploiting the structure of temperature series. We reduce the number of distances to be sorted by only looking for neighbours among days that are likely to be similar to  $t$  — summer days tend to be similar to summer days, and so on. This idea is the basis of the *Seasonal Window*.

To use this method, the window half-width  $w$  must first be specified. The potential neighbours of  $\mathbf{v}_t$  are the observations which are within  $w$  days of  $\mathbf{v}_{[t+cT]}$ , where  $T$  is the period,  $c = \dots, -1, 0, 1, \dots$ . For example, if  $w = 10$  and day  $t$  falls on June 17th, we would look at June 7th to 27th for each year. The method then proceeds the same way as simple  $k$ -NN.

### The Threshold Method

We also propose another approach to the problem, which is to simplify the search by allowing the size of the neighbourhood to vary. We term this the *Threshold method*. To begin with, we need to specify acceptable lower and upper bounds for the number of points in the neighbourhood of  $\mathbf{v}_t$ . Denote these as  $k_l$  and  $k_u$  respectively. Then, for each  $\mathbf{v}_t$ , all the distances are calculated and the points  $\mathbf{v}_s$  with distance from  $\mathbf{v}_t$  smaller than some given *threshold*  $r_t$  ( $d(\mathbf{v}_t, \mathbf{v}_s) < r_t$ ) are included in the test neighbourhood of  $\mathbf{v}_t$ . Let the size of this neighbourhood be  $k'$ . If  $k' < k_l$  we need to raise the threshold  $r_t$  by a small amount; if  $k' > k_u$ , we need to lower it. If  $k' \in [k_l, k_u]$ , then the threshold is deemed acceptable and the  $k'$ -NN estimate is calculated. We then set  $r_{t+1} = r_t$  and proceed the same way for  $\mathbf{v}_{t+1}$ . The first threshold is set arbitrarily.

The main computational advantage of this method is that, because of the similarity of consecutive daily temperatures, if  $k_l$  and  $k_u$  are far enough apart, then the threshold will not need to be adjusted very often — most of the time, an acceptable value for  $r_t$  will also be an acceptable value for  $r_{t+1}$ . This tends to make up for the cost of calculating all the distances. Clearly, the larger the difference between  $k_l$  and  $k_u$  the faster this process will be. Another advantage is that we are likely to pick larger neighbourhoods for points with many close neighbours and smaller ones for points with few close neighbours, which is a more natural way of dealing with the bias-variance trade-off.

## 6.3 Modelling Daily Temperature Series

### 6.3.1 Motivation

In the ad hoc methods described so far, we make no attempt to model daily temperatures directly, apart from accounting for fixed-cycle periodic changes in the daily mean and variance. We will now approach the problem of missing-temperature estimation in a different way. We will first model the deseasonalised daily temperatures  $\{z_{i,t}\}$  directly.

In daily temperature series, the sample autocorrelations tend to be significantly different from 0 even for very large lags (Figure A.5), which means that these series can not be adequately modelled with a small number of ARMA parameters. For this reason, we choose to apply a long-memory model to the series.

We demonstrate that it is possible to filter the deseasonalised series in a way that eliminates the long-memory behaviour, but preserves the short-term autocorrelation structure. Subsequently, we will approach the resulting series using the various autoregressive models described in the technical introduction. These models can then be cast in state-space form, which enables us to make use of the UR representation (Chapter 4), as well as the filtering, smoothing and interpolation results (Chapter 3) to improve the accuracy of the estimates.

### 6.3.2 The ARFIMA Model

For the series  $\{z_{i,t}\}$ , the autoregressive fractionally integrated moving average (ARFIMA) model is defined as:

$$\Phi(B)(1-B)^d(z_{i,t} - \nu_{i,t}) = \Theta(B)\epsilon_{i,t}, \quad t = 1, \dots, n \quad (6.51)$$

where

$$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (6.52)$$

is the autoregressive polynomial,

$$\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (6.53)$$

is the moving average polynomial,  $\nu_{i,t}$  is the mean of  $z_{i,t}$ ,  $B$  is the backward shift operator,  $p, q$  are non-negative integers, and the errors are a white noise process ( $\{\nu_{i,t}\} \sim \text{WN}(0, \sigma_i^2)$ ). We require that the ARMA part of the model is stationary and invertible, which is equivalent to requiring that the roots of  $\Phi(x) = 0$  and  $\Theta(x) = 0$  lie outside the unit circle. We assume that the polynomials  $\Phi(x)$  and  $\Theta(x)$  have no common factors — if they do, a model with identical properties can be constructed by reducing both  $p$  and  $q$  by one.

The model equation (6.51) is identical to that of an ARIMA( $p, d, q$ ) model except that in the ARFIMA model we allow the differencing parameter  $d$  to take any real value. The fractional difference operator  $(1 - B)^d$  is defined by the binomial expansion

$$(1 - B)^d = 1 + \sum_{k=1}^{\infty} \frac{\Gamma(k - d)}{\Gamma(-d)\Gamma(k + 1)} B^k \quad (6.54)$$

where  $\Gamma(r)$  is the gamma function,  $\Gamma(r) = (r - 1)!$  if  $r$  is a positive integer. Using the property  $\Gamma(r) = (r - 1)\Gamma(r - 1)$ , the expansion can be written as:

$$\begin{aligned}
(1 - B)^d &= 1 + \sum_{k=1}^{\infty} \frac{(k - d - 1)(k - d - 2) \dots (k - d - k) \Gamma(-d)}{\Gamma(-d) k!} B^k \\
&= 1 + \sum_{k=1}^{\infty} \frac{(k - d - 1)(k - d - 2) \dots (1 - d)(-d)}{k!} B^k \\
&= 1 - dB - \frac{1}{2}d(1 - d)B^2 - \frac{1}{6}d(1 - d)(2 - d)B^3 - \dots
\end{aligned} \tag{6.55}$$

One of the key properties of time series that display long-memory behaviour is the slow decay of the correlations. The slow decay of the coefficients in expansion (6.55) allows ARFIMA to model long memory processes such as daily temperature series while only using a small number of parameters.

### 6.3.3 Estimation of $d$ - Aggregated Variance

Hosking (1981) proves that the autocorrelation function  $\rho(\tau)$  of an ARFIMA( $p, d, q$ ) process decreases hyperbolically if  $d \in (0, 1)$ , that is,

$$\rho(\tau) \approx C\tau^{2d-1} \quad \text{as } \tau \rightarrow \infty \tag{6.56}$$

where  $C$  is a constant. This is considerably slower than the standard ARMA case, where the correlations decay exponentially after the largest lag of the MA component. Beran (1989) shows that, if the correlations of a stationary process decay like  $\tau^{2H-2}$  (where  $H \equiv d + \frac{1}{2} \in (0, 1)$  is known as the Hurst parameter), then the variance of an  $n$ -member sample mean will decay like  $n^{2H-2}$ , rather than  $n^{-1}$ . The ARFIMA process is stationary for  $d < \frac{1}{2}$  so, given  $n$  consecutive observations  $z_{i,1}, z_{i,2}, \dots, z_{i,n}$ , we have

$$\text{Var}\left(\frac{1}{n} \sum_{i=1}^n z_{i,t}\right) \approx Cn^{2d-1} \quad \text{as } n \rightarrow \infty \quad (6.57)$$

This result can be used to estimate  $d$  in long data series. The algorithm proposed by Beran (1994) consists of the following steps:

1. Divide the series into  $k = n/m$  blocks of size  $m$
2. Compute  $\bar{z}_{i,k(m)}$ , the  $m$ -member mean of each block

$$\bar{z}_{i,k(m)} = \frac{1}{m} \sum_{t=(k-1)m+1}^{km} z_{i,t} \quad (6.58)$$

3. Estimate  $s^2(m)$ , the variance of the block mean ( $\bar{z}_i$  is the overall mean)

$$s^2(m) = \frac{1}{n/m - 1} \sum_{k=1}^{n/m} (\bar{z}_{i,k(m)} - \bar{z}_i)^2 \quad (6.59)$$

4. Repeat for several values of  $m$  and then regress  $\log s^2(m)$  on  $\log m$  — the slope of the least-squares line will yield an estimate of  $2d-1$ , and hence an estimate of  $d$ .

If the series does not display long-memory behaviour, the slope of the log-log line will be approximately equal to  $-1$  (corresponding to  $d \approx 0$ ). This procedure is known as the *aggregated variance* (AV) method (Figure A.6). If the data are relatively homogeneous (as we would expect for temperature data), the AV estimate of  $d$  tends to be close to the maximum likelihood (ML) estimate. The advantage is that the AV method involves a much smaller number of computations, which makes it significantly faster.

In practice, the choice of upper and lower bounds for  $m$  is crucial. If the value of  $m$  is low, then the blocks used to estimate the sample mean variance are very small, which can introduce bias due to short range



effects. On the other hand, if the value of  $m$  is very large, then the small number of blocks ( $n/m$ ) will make the estimate of  $s^2(m)$  unstable (Park *et al.*, 2004).

#### 6.3.4 Differenced Variance

Another disadvantage of the AV method, as well as other heuristic estimators, is that it will occasionally produce a positive estimate for  $d$  even in the absence of long memory (Bhattacharya *et al.*, 1983). Series that contain slowly decaying trends or shifts in the mean are most susceptible to this problem.

An improved method for estimating  $d$  was proposed by Teverovsky and Taqqu (1997). It involves fitting a straight line to a log-log plot of the first-order differences of  $s^2(m_j)$  versus  $m_j$ , where  $m_1, m_2, \dots$  are logarithmically spaced.

For any smooth function  $f(x)$  in a small interval  $(x_1, x_2)$ , we have

$$\begin{aligned} \left. \frac{df(x)}{dx} \right|_{x=x_1} &\approx \frac{f(x_2) - f(x_1)}{x_2 - x_1} \\ \Leftrightarrow f(x_2) - f(x_1) &\approx \left. \frac{df(x)}{dx} \right|_{x=x_1} (x_2 - x_1) \end{aligned} \quad (6.60)$$

We now apply (6.60) to the function  $s^2(m)$  on the interval  $(m_j, m_{j+1})$ , and take logarithms on both sides:

$$\log(s^2(m_{j+1}) - s^2(m_j)) \approx \log\left(\left. \frac{ds^2(m)}{dm} \right|_{m=m_j}\right) + \log(m_{j+1} - m_j) \quad (6.61)$$

We know that  $s^2(m) \approx Cm^{2d-1}$  for some constant  $C$ , so:

$$\left. \frac{ds^2(m)}{dm} \right|_{m=m_j} \approx (2d-1)Cm_j^{2d-2} \quad (6.62)$$

Since the points are logarithmically spaced, the ratio between consecutive points is equal to some constant  $C_1$ . Thus, the second term on the RHS of (6.61) yields:

$$\begin{aligned} \log(m_{j+1} - m_j) &= \log(C_1 m_j - m_j) \\ &= \log\{m_j(C_1 - 1)\} \\ &= \log m_j + \log C_2 \end{aligned} \quad (6.63)$$

where  $C_2$  is also a constant. Hence

$$\begin{aligned} \log(s^2(m_{j+1}) - s^2(m_j)) &\approx \log\{(2d-1)C\} + (2d-2)\log m_j \\ &\quad + \log m_j + \log C_2 \\ \Rightarrow \log(s^2(m_{j+1}) - s^2(m_j)) &\approx (2d-1)\log m_j + C_3 \end{aligned} \quad (6.64)$$

for some constant  $C_3$ . Thus, a log-log plot of the first-order differences  $s^2(m_{j+1}) - s^2(m_j)$  versus  $m_j$  will also produce a straight line with slope  $2d-1$ . This is known as the differenced variance (DV) method. Differencing the series increases scatter (Montanari *et al.*, 1997) so it is necessary to fit the line using a method more robust than least squares, such as LAD. The points corresponding to a negative difference of variances must be ignored, as they have no logarithm.

### 6.3.5 Maximum-Likelihood Methods

The methods used for estimating  $d$  are heuristic; they are straight-forward to apply and intuitive, since they take advantage of the key properties of long-memory series. They have the additional advantage of not requiring us to place distributional assumptions on the error terms. Finally, they allow us to estimate  $d$  independently of the other model parameters. However, it is far from easy to establish the theoretical properties of the resulting estimators or use them to construct confidence intervals for  $d$ .

Exact maximum-likelihood estimation (EML) involves estimating  $d$  jointly with the AR and MA parameters. If we assume that the errors are Gaussian, the p.d.f. of  $\mathbf{Z}_i = (z_{i,1}, z_{i,2}, \dots, z_{i,n})$  is

$$f(\mathbf{Z}, \mathbf{S}) = (2\pi)^{-n/2} |\mathbf{S}|^{1/2} \exp\left(-\frac{1}{2} \mathbf{Z}' \mathbf{S}^{-1} \mathbf{Z}\right) \quad (6.65)$$

where  $\mathbf{S}$  is the  $n \times n$  covariance matrix.  $z_{i,t}$  is stationary, so

$$\mathbf{S} = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(2) & \cdots & \gamma(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix} \quad (6.66)$$

The log-likelihood is

$$\ell(\phi, \theta, d) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{S}| - \frac{1}{2} \mathbf{Z}' \mathbf{S}^{-1} \mathbf{Z} \quad (6.67)$$

where  $\phi = (\phi_1, \dots, \phi_p)$  and  $\theta = (\theta_1, \dots, \theta_q)$ . The matrix  $\mathbf{S}$  is a function of the autoregressive, moving average and fractional differencing parameters. The core of the EML approach is the computation of  $\mathbf{S}$  as a function of these parameters (Doornik & Ooms, 2004). EML is often prohibitively expensive from a computational viewpoint, but there are a

number of approximate maximum-likelihood methods that can be used instead.

### 6.3.6 Removing Long Memory

Because the deseasonalised series  $\{z_{i,t}\}$  has zero mean, the ARFIMA model equation can be written in the form

$$(1 - B)^d z_{i,t} = w_{i,t}, \quad t = 1, \dots, n \quad (6.68)$$

where the filtered series  $\{w_{i,t}\}$  is stationary. Fitting an ARFIMA( $p, d, q$ ) model to the series  $\{z_{i,t}\}$  is equivalent to fitting an ARMA( $p, q$ ) model to the filtered series.  $\{w_{i,t}\}$  can be obtained using the equation

$$\begin{aligned} w_{i,t} &= (1 - B)^d z_{i,t} \\ &= \left\{ 1 - dB - \frac{1}{2}d(1-d)B^2 - \dots \right\} z_{i,t} \\ &= z_{i,t} - dz_{i,t-1} - \frac{1}{2}d(1-d)z_{i,t-2} - \dots \end{aligned} \quad (6.69)$$

This approach is successful in removing the long-term dependence from temperature series, in the sense that there is no evidence of long memory in the filtered series. However, the first two sample partial autocorrelations of  $\{w_{i,t}\}$  tend to be high, because of the short-term structure in temperature series (Figure A.7). This needs to be modelled separately, with the inclusion of an autoregressive component.

We consider three different autoregressive models: a univariate AR( $p$ ) model for each temperature series  $\{w_{i,t}\}$  (note that this is equivalent to fitting a ARFIMA( $p, d, 0$ ) model to  $\{z_{i,t}\}$ ); a system of seemingly-unrelated autoregressions (SUAR) for the series  $\{w_{1,t}\}, \dots, \{w_{p,t}\}$ ; and a full vector autoregression (VAR) for the multivariate series  $\{\mathbf{w}_t\} =$

$\{(w_{1,t}, \dots, w_{p,t})\}$ . These models allow for different degrees of interdependence between temperature series at neighbouring locations.

### 6.3.7 Model-based Estimation

#### General Approach

The autoregressive models can be expressed in state space form, so in theory we could apply the Kalman filter smoother recursions to obtain an interpolated estimate of observation  $w_{i,t}$ , conditional on the whole sample. Let this be  $\hat{w}_{i,t}$ . Note that this is the interpolated estimate obtained by removing a single component of the vector  $\mathbf{w}_t$ , thus it is essential to use the UR of the series.

The long-memory structure can then be reconstructed using the equation

$$\begin{aligned}
 w_{i,t} &= (1 - B)^d z_{i,t} \\
 \Leftrightarrow z_{i,t} &= (1 - B)^{-d} w_{i,t} \\
 &= \left\{ 1 + dB + \frac{1}{2}d(1+d)B^2 + \dots \right\} w_{i,t} \\
 &= w_{i,t} + dw_{i,t-1} + \frac{1}{2}d(1+d)w_{i,t-2} + \dots \quad (6.70)
 \end{aligned}$$

Let  $\hat{z}_{i,t}$  be the estimate of the temperature departure obtained from equation (6.70) if we substitute  $\hat{w}_{i,t}$  for  $w_{i,t}$ . The missing temperature  $y_{i,t}$  will thus be estimated from the estimated departure as

$$\hat{y}_{i,t} = \hat{\mu}_{i,t} + \hat{\sigma}_{i,t} \hat{z}_{i,t} \quad (6.71)$$

in direct analogy to the harmonic departures method.

## The Algorithm

We soon encounter a practical problem in the application of this method. If observation  $y_{i,t}$  is missing, then the departure  $z_{i,t}$  will be unknown. In the filtered series, the terms  $w_{i,t}, w_{i,t+1}, \dots, w_{i,n}$  depend on  $z_{i,t}$  hence they will also be unknown. This makes it impossible to produce the filtered series and hence apply the smoothing and interpolation results directly.

To overcome this problem, we propose the following algorithm:

1. Estimate the long-memory parameter  $d$  using the DV method, ignoring the blocks which include the missing value  $z_{i,t}$ .
2. Filter the deseasonalised series up to time  $t - 1$  to remove the long-memory structure and obtain terms  $w_{i,1}, \dots, w_{i,t-1}$ .
3. Fit an AR(2) model to  $w_{i,1}, \dots, w_{i,t-1}$ , estimating the autoregressive parameters using least-squares regression, and express the model in state-space form.
4. Compute the initial estimate  $\dot{w}_{i,t}^{(0)} = L(w_{i,t} | w_{i,1}, \dots, w_{i,t-1})$  using the Kalman filter recursions.
5. Estimate the remaining terms of the filtered series  $w_{i,t+1}, \dots, w_{i,n}$  using expression (6.69), replacing  $w_{i,t}$  with its estimate  $\dot{w}_{i,t}^{(j)}$ . Let the new series be  $w_{i,1}, \dots, w_{i,t-1}, \dot{w}_{i,t}^{(j)}, \dots, \dot{w}_{i,n}^{(j)}$ .
6. Update the estimates of the autoregressive parameters using the new series, and apply the UR KFS recursions to produce the interpolated estimate  $\dot{w}_{i,t}^{(j+1)} = L(w_{i,t} | w_{i,1}, \dots, w_{i,t-1}, \dot{w}_{i,t+1}^{(j)}, \dots, \dot{w}_{i,n}^{(j)})$ .
7. Update the autoregressive parameters by fitting an AR(2) model to  $\{w_{i,1}, \dots, w_{i,t-1}, \dot{w}_{i,t}^{(j+1)}, \dot{w}_{i,t+1}^{(j)}, \dots, \dot{w}_{i,n}^{(j)}\}$ .

Steps 5-7 need to be repeated until the estimate converges. In practice, it was found that a single iteration of the algorithm is sufficient: subsequent estimates  $w_{i,t}^{(2)}, w_{i,t}^{(3)}, \dots$  do not differ greatly from the first interpolated estimate  $w_{i,t}^{(1)}$ .

## SUAR and VAR models

If the chosen autoregressive model is multivariate, we need to make a few slight modifications to the algorithm. We will use the univariate representation of the multivariate state space model, as described in Chapter 4. The following changes are required:

- In step 4, we can compute the contemporaneous terms of the filtered series  $\mathbf{w}_{t \setminus i} = (w_{1,t}, \dots, w_{i-1,t}, w_{i+1,t}, \dots, w_{i,t})'$ , thus we can replace the initial estimate of  $w_{i,t}$  with the partially interpolated estimate (Chapter 3)

$$\dot{w}_{i,t}^{(0)} = E(w_{i,t} | \mathbf{w}_1, \dots, \mathbf{w}_{t-1}, \mathbf{w}_{t \setminus i}). \quad (6.72)$$

- In step 5, the terms  $\mathbf{w}_{t+1 \setminus i}, \dots, \mathbf{w}_{n \setminus i}$  do not depend on  $w_{i,t}$  so they can be computed directly
- In step 6, the updated value  $\dot{w}_{i,t}^{(j+1)}$  is the smoothed estimate

$$\dot{w}_{i,t}^{(j+1)} = E(w_{i,t} | \mathbf{w}_1, \dots, \mathbf{w}_{t-1}, \mathbf{w}_{t \setminus i}, \dots, \mathbf{w}_{n \setminus i}, \dot{w}_{i,t+1}^{(j)}, \dots, \dot{w}_{i,n}^{(j)}) \quad (6.73)$$

which is produced by the univariate representation of the Kalman filter smoother.

The UR Kalman filter converges periodically to a steady state very quickly (Chapter 5), which drastically reduces the volume of data to be stored during the forward pass. We only require the output matrices  $\tilde{P}_{j,s}, \tilde{K}_{j,s}, \tilde{F}_{j,s}$  until periodic convergence is achieved.

## 6.4 Results and Conclusions

### 6.4.1 Station selection

A standard approach for assessing the accuracy of the different methods is to take a serially complete temperature series, treat each observation as missing and estimate it from the rest of the data. This estimate can then be compared to the true temperature on that day.

The data used in this application are daily maximum temperatures from January 1950 to December 2001 at locations ELP (El Paso, TX) and LBB (Lubbock, TX). These locations are displayed in Figure A.8.

The climate in Texas varies widely, from arid to wet. Extreme weather phenomena such as tornadoes, hurricanes and thunderstorms can cause violent sudden changes in temperature, particularly in the northernmost regions. Additionally, Texas produces more greenhouse gases than any other U.S. state, and is affected by climate change as a result.

The choice of locations is deliberate and was designed to illustrate how the models behave under significantly different circumstances. LBB is closer to the northwestern part of Texas, where the station density is high the nearest weather station, AMA (Amarillo, TX), is around 100 miles away and the two temperature series are very highly correlated ( $r = 0.96$ ). On the other hand, ELP is the westernmost station in the record and is significantly further from its neighbours. The weather stations included in the model for LBB, chosen according to the selection criteria recommended by DeGaetano *et al.* (1995), were ABI (Abilene, TX), AMA (Amarillo, TX) and MAF (Midland, TX). Similarly, the neighbours of ELP were taken to be AMA, LBB and MAF.

Monthly summary statistics for the two main series can be seen in Table B.1. The series AMA is also included in this table, to illustrate the difference in temperature between the Northern Plains region of the state, where AMA is located, and the more arid western region (ELP). Series ELP displays considerable lower variability within each month compared to the other two series.

### 6.4.2 Ad-hoc Methods

#### Comparison of Existing Methods

We begin by comparing the performance of the Departures and LAD methods. In the following tables, ‘mean’ and ‘med’ refer to the the mean and median of the errors  $y_{i,t} - \hat{y}_{i,t}$ ; MAE and mdAE are, respectively, the mean and the median of the absolute errors  $|y_{i,t} - \hat{y}_{i,t}|$ ; sd and sdA are the estimated standard deviations of the errors and the absolute errors,



respectively, while MSE is the mean square error  $\frac{1}{n} \sum_t (y_{i,t} - \hat{y}_{i,t})^2$ . As a baseline, we also provide summary statistics for the following simple models: Mean ( $\hat{y}_{i,t} = \hat{\mu}_{i,t}$ , the daily mean) and 2pMA ( $\hat{y}_{i,t} = (y_{i,t-1} + y_{i,t+1})/2$ , a 2-point moving average)

As can be seen in Tables B.2 and B.3, the LAD approach is slightly more biased than the departure approach in terms of the mean error, but performs better in almost every other category (errors have units of degrees Fahrenheit). These results confirm that the LAD method achieves much smaller mean and median absolute errors than the departure method and also that its estimates have smaller variance. As would be expected, both techniques produce far more accurate results for LBB than for ELP, due to the proximity of its neighbouring stations. Note that this is not affected by the higher variance of the daily temperatures at LBB.

## Harmonic Departures

We will now assess the accuracy of the harmonic departures method, denoted by ‘Harm’. Tables B.4 and B.5 demonstrate that, despite being based on the same principle as De Gaetano’s departure method, Harmonic Departures are noticeably more accurate. In fact, for LBB the Harmonic Departure model produces error statistics comparable to those from the LAD method. This is a significant result because between-station methods tend to be computationally efficient compared with LAD, which involves calculating a set of regression coefficients, and hence solving a quadratic programming problem, for each month.

## $k$ -Nearest Neighbours and Approximations

The results from the models in this section are summarised in Tables B.6 and B.7. Two of the previous models — Harmonic Departures and LAD — are also included for ease of comparison. In the following table NN( $k$ ) denotes the basic  $k$ -Nearest Neighbours model; T( $k_l, k_u$ ) denotes the Threshold method, with neighbourhood size lower bound  $k_l$  and upper bound  $k_u$ ; and W( $k, w$ ) is the Seasonal Window, with neighbourhood size  $k$  and window half-width  $w$ .

All of these methods result in similar error statistics; they generally outperform Harmonic Departures by a substantial margin and are comparable to LAD, the best model so far. Surprisingly, the Seasonal Window method occasionally does a little bit better than  $k$ -NN, the method it is approximating. This needs to be discussed in more detail.

## Seasonal Window Results

The Seasonal Window often results in better error statistics than simple  $k$ -NN, even though it only uses a small subset of the original training sample to construct the estimates. This phenomenon can be explained through the structure of the data: say we are trying to estimate the temperature on  $y_{i,t}$ , a summer day, at location ELP and the  $j^{th}$  predictor variable is location AMA. If the weather at AMA happens to be unseasonably cold on that day,  $y_{j,t}$  could easily be closer to, say, a relatively warm spring day  $y_{j,s}$  rather than other summer days. If the number of predictors is small, the effect of this on the overall distance metric could be enough to lead to  $v_s$  being included in the neighbourhood of  $v_t$ .

However, it is a well-documented fact (e.g. Allen & DeGaetano, 2001) that, given the temperature at AMA is extreme, it is still much likelier that the temperature at nearby ELP will not be extreme. Thus, the difference between  $y_{i,t}$  and  $y_{i,s}$  will potentially be large, resulting in a larger prediction error. With the Seasonal Window method,  $y_{i,s}$  would automatically be excluded from the estimation process and, more often than not, replaced by days with temperatures closer to  $y_{i,t}$ .

### 6.4.3 Model-based Methods

We now compare the estimates produced by the model-based autoregressive methods. Only the autoregressive coefficients corresponding to the first two lags are significantly different from zero. Thus, the models fitted are AR(2), SUAR(2) and VAR(2). Tables B.8 and B.9 contain summary statistics of the residuals resulting from a single iteration of the algorithm.

By the classification given in 4.2.1, the AR(2) model is a within-station method and the estimates it produces are, unsurprisingly, substantially

less accurate than those from the multivariate autoregressive models. Note that, in terms of its residuals, it is virtually indistinguishable from a simple 2-point moving average, which does not take into account the long-memory structure. We can also see that the SUAR- and VAR-based estimation methods are comparable to the most accurate ad-hoc approach, LAD.

The numerical results in this section were generated using programs written in Ox, an object-oriented matrix programming language with a comprehensive mathematical and statistical function library (Doornik, 2002). Full details and program code can be found on the author's thesis page — <http://stats.lse.ac.uk/milt/phd>

## Appendix A

### Figures

Figure A.1: Periodogram of  $\hat{y}_{i,t}$  (Lubbock, TX)

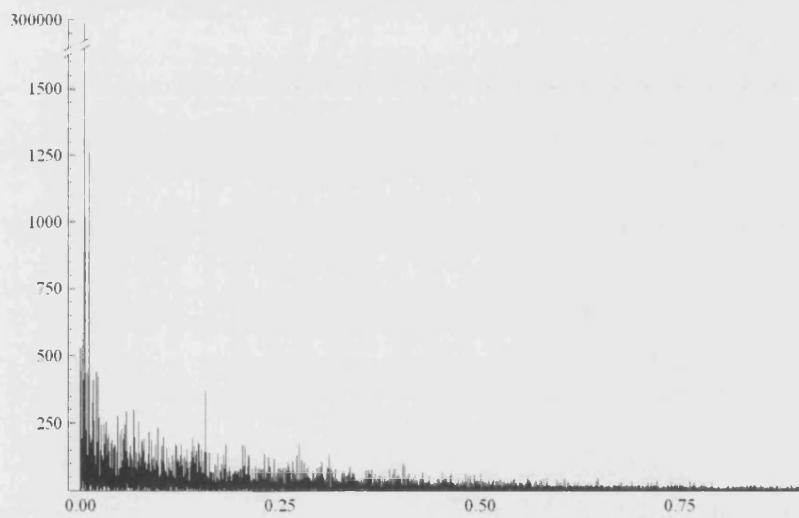


Figure A.2: Periodogram of  $(\zeta_{i,t})$  (Lubbock, TX)

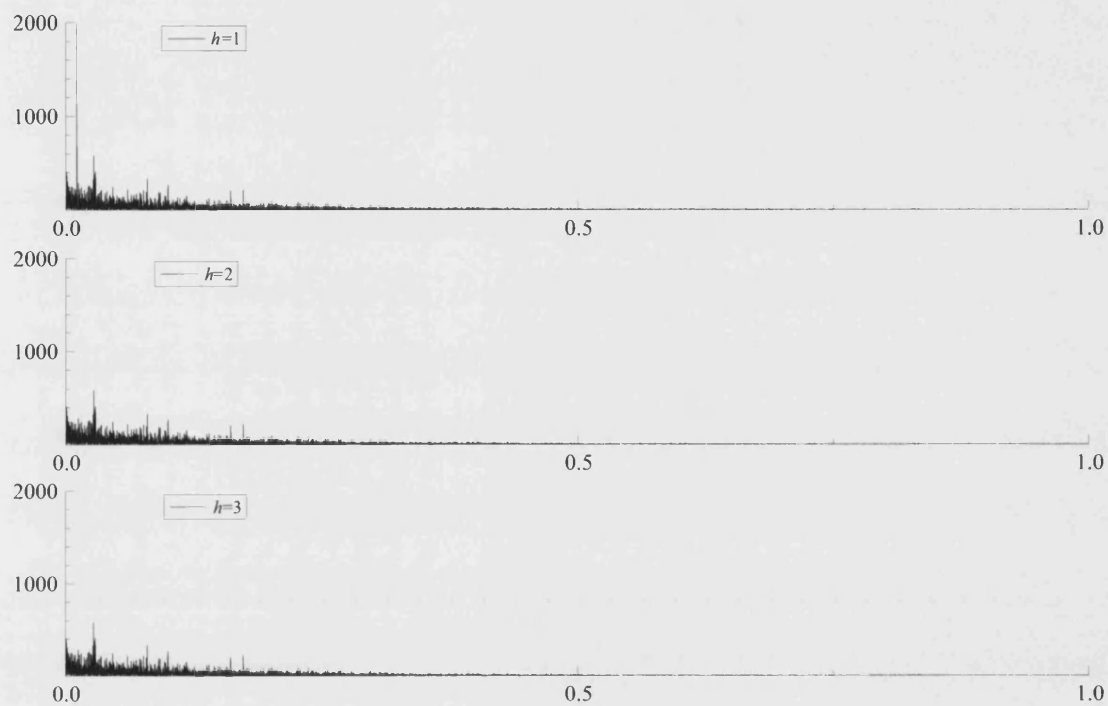


Figure A.3: Periodogram of  $(\zeta_{i,t}^*)^2$  (Lubbock, TX)

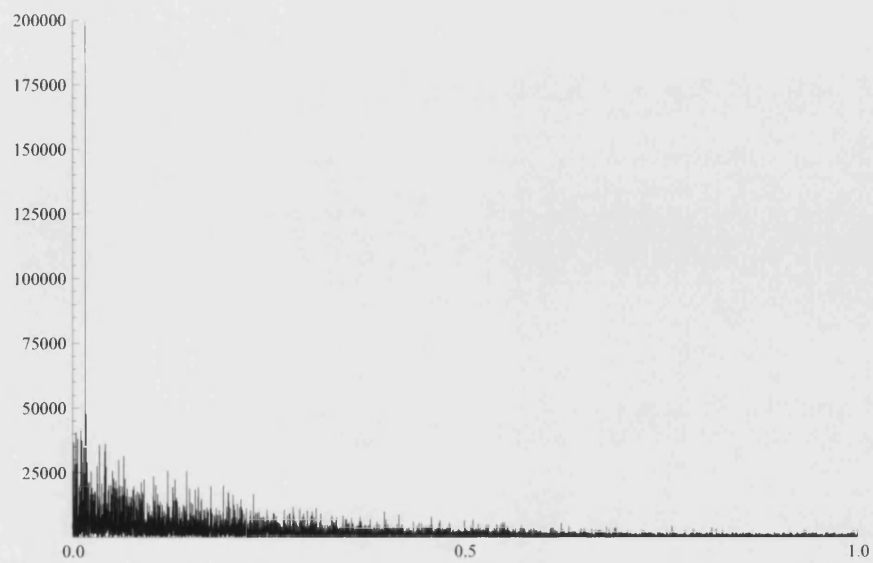


Figure A.4: Periodogram of  $z_{i,t}$  (Lubbock, TX)

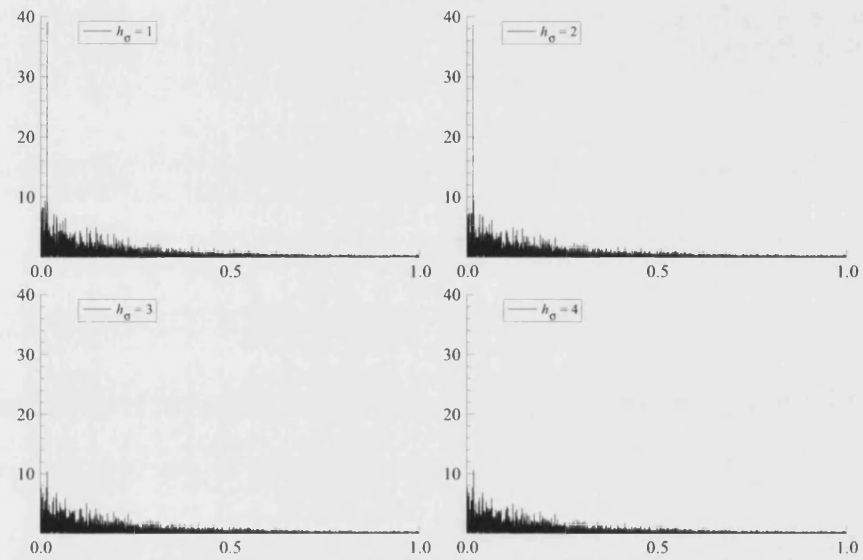


Figure A.5: ACF of the deseasonalised series (Lubbock, TX)

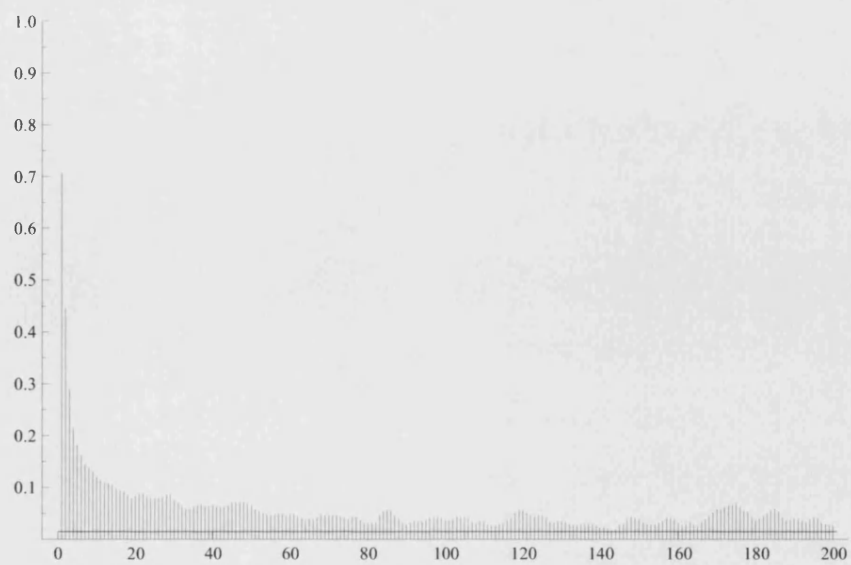


Figure A.6: Aggregated Variance log-log plot (Lubbock, TX)

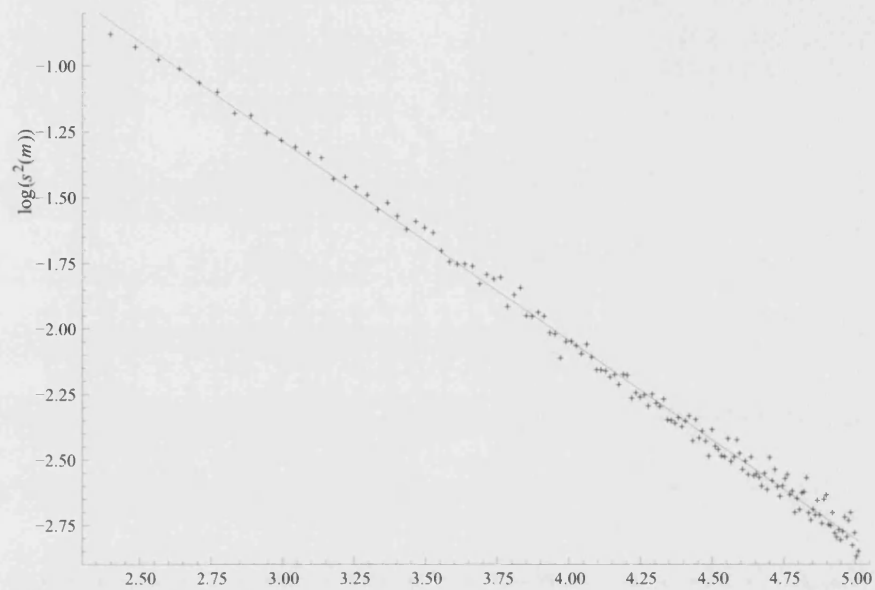


Figure A.7: Partial ACF of  $\{w_{i,t}\}$  (Lubbock, TX)

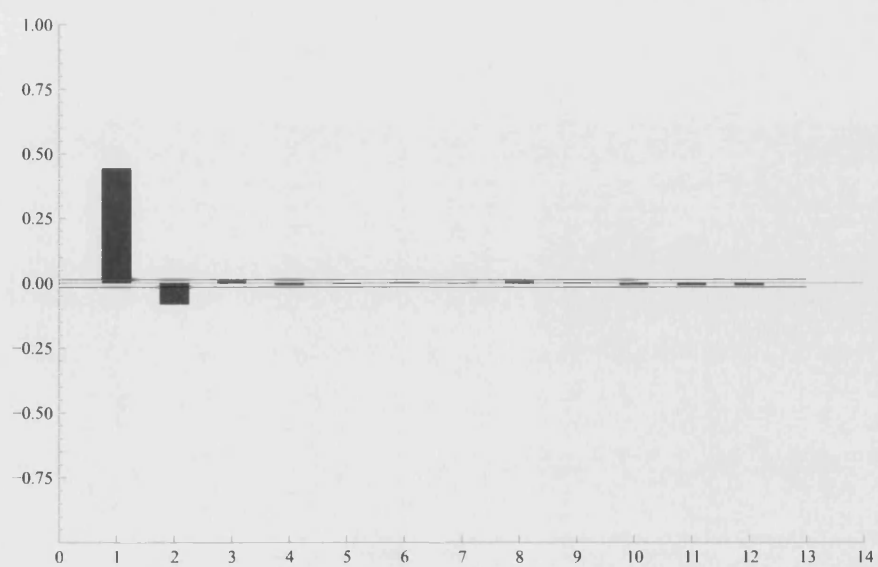
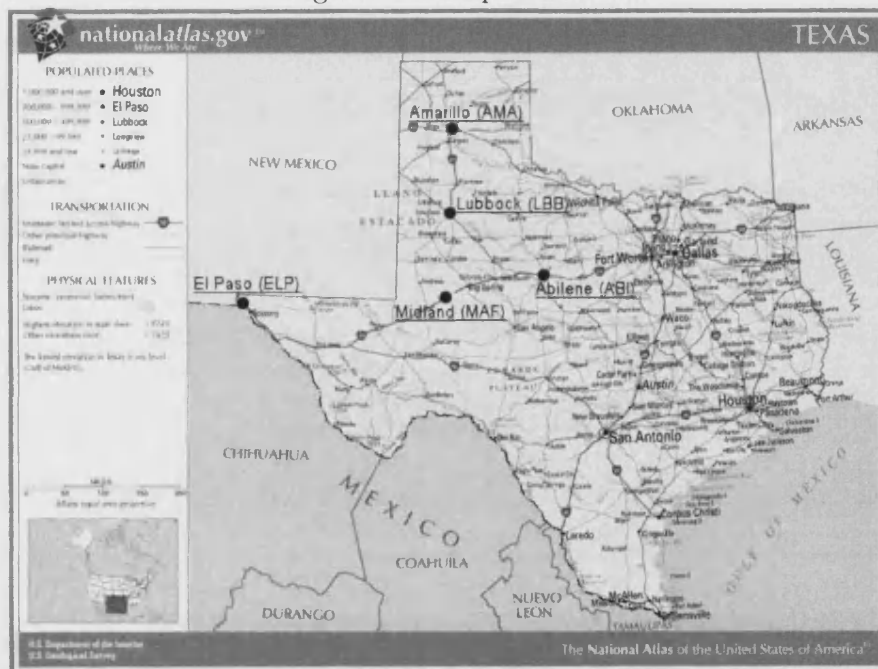


Figure A.8: Map of Texas





## Appendix B

### Tables

Table B.1: Summary statistics (El Paso, TX and Lubbock, TX)

	ELP mean	ELP s.d.	LBB mean	LBB s.d.	AMA mean	AMA s.d.
Jan	57.94	8.73	53.34	13.55	49.50	14.13
Feb	63.38	8.91	58.15	13.58	53.85	14.38
Mar	69.95	8.77	65.62	12.51	61.32	13.45
Apr	78.69	8.07	75.09	10.87	70.98	11.59
May	87.39	6.82	83.17	9.57	79.16	10.23
Jun	95.87	6.10	90.62	7.88	87.87	8.42
Jul	95.24	5.48	92.43	6.24	91.36	6.51
Aug	93.03	5.31	90.58	6.25	89.42	6.81
Sep	87.92	6.91	83.87	8.96	82.29	9.78
Oct	78.75	8.14	75.02	10.27	72.48	11.24
Nov	66.11	8.94	62.55	11.97	59.08	12.80
Dec	58.08	8.64	54.66	12.55	50.78	13.26
Overall	77.75	15.70	73.83	17.53	70.75	18.67

Table B.2: Results from existing estimation methods (El Paso, TX)

Model	mean	MAE	med	mdAE	sd	sdA	MSE
Mean	0.000	5.653	0.850	4.616	7.319	4.648	53.561
2pMA	-0.001	3.204	0.000	2.000	4.441	3.075	19.720
Dep	-0.016	4.012	-0.233	3.173	5.271	3.419	27.786
LAD	0.031	2.945	0.000	2.263	3.957	2.643	15.657

Table B.3: Results from existing estimation methods (Lubbock, TX)

Model	mean	MAE	med	mdAE	sd	sdA	MSE
Mean	0.000	8.067	1.165	6.486	10.400	6.564	108.170
2pMA	0.000	5.173	0.500	3.500	7.118	4.890	50.670
Dep	0.015	3.279	0.111	2.611	4.257	2.714	18.118
LAD	-0.061	2.237	0.000	1.684	3.040	2.060	9.245

Table B.4: Results from Harmonic Departures methods (El Paso, TX)

Model	mean	MAE	med	mdAE	sd	sdA	MSE
Dep	-0.016	4.012	-0.233	3.173	5.271	3.419	27.786
Harm	-0.002	3.401	-0.162	2.641	4.535	3.000	20.568
LAD	0.031	2.945	0.000	2.263	3.957	2.643	15.657

Table B.5: Results from Harmonic Departures methods (Lubbock, TX)

Model	mean	MAE	med	mdAE	sd	sdA	MSE
Dep	0.015	3.279	0.111	2.611	4.257	2.714	18.118
Harm	0.000	2.440	0.172	1.875	3.267	2.173	10.673
LAD	-0.061	2.237	0.000	1.684	3.040	2.060	9.245

Table B.6: Results from NN Methods (El Paso, TX)

Model	mean	MAE	med	mdAE	sd	sdA	MSE
Harm	-0.002	3.401	-0.162	2.641	4.535	3.000	20.568
LAD	0.031	2.945	0.000	2.263	3.957	2.643	15.657
NN(10)	0.041	3.071	0.050	2.389	4.095	2.709	16.770
NN(50)	0.088	3.011	0.082	2.327	4.022	2.668	16.182
T(1,100)	0.031	3.153	0.000	2.412	4.240	2.835	17.981
T(5,200)	0.069	3.027	0.050	2.333	4.047	2.688	16.386
T(10,100)	0.071	3.029	0.075	2.348	4.044	2.680	16.359
W(10,5)	0.078	3.083	0.200	2.400	4.126	2.744	17.033
W(10,15)	0.060	3.008	0.100	2.300	4.037	2.693	16.299
W(10,30)	0.072	2.995	0.100	2.300	4.021	2.683	16.173

Table B.7: Results from NN Methods (Lubbock, TX)

Model	mean	MAE	med	mdAE	sd	sdA	MSE
Harm	0.000	2.440	0.172	1.875	3.267	2.173	10.673
LAD	-0.061	2.237	0.000	1.684	3.040	2.060	9.245
NN(10)	-0.012	2.355	0.071	1.813	3.172	2.124	10.061
NN(50)	-0.027	2.294	0.061	1.755	3.094	2.076	9.575
T(1,100)	-0.018	2.421	0.050	1.846	3.302	2.246	10.906
T(5,200)	-0.018	2.315	0.083	1.781	3.124	2.098	9.762
T(10,100)	-0.025	2.310	0.065	1.767	3.114	2.089	9.699
W(10,5)	-0.080	2.395	0.000	1.800	3.223	2.159	10.396
W(10,15)	-0.041	2.329	0.000	1.800	3.146	2.115	9.900
W(10,30)	-0.025	2.315	0.100	1.700	3.123	2.096	9.752

Table B.8: Results from model-based methods (El Paso, TX)

Model	mean	MAE	med	mdAE	sd	sdA	MSE
2pMA	-0.001	3.204	0.000	2.000	4.441	3.075	19.720
Harm	-0.002	3.401	-0.162	2.641	4.535	3.000	20.568
LAD	0.031	2.945	0.000	2.263	3.957	2.643	15.657
AR	-0.001	3.174	0.358	2.259	4.375	3.010	19.138
SUAR	0.000	2.445	0.063	1.798	3.330	2.261	11.090
VAR	0.000	2.375	0.090	1.767	3.243	2.208	10.520

Table B.9: Results from model-based methods (Lubbock, TX)

Model	mean	MAE	med	mdAE	sd	sdA	MSE
2pMA	0.000	5.173	0.500	3.500	7.118	4.890	50.670
Harm	0.000	2.440	0.172	1.875	3.267	2.173	10.673
LAD	-0.061	2.237	0.000	1.684	3.040	2.060	9.245
AR	0.000	5.154	0.469	3.717	7.017	4.761	49.239
SUAR	-0.001	2.034	0.033	1.462	2.832	1.971	8.020
VAR	-0.001	1.998	0.049	1.438	2.785	1.940	7.757

# Bibliography

- [1] Allen, R. J. and DeGaetano, A. T., 2001, Estimating Missing Daily Temperature Extremes Using an Optimized Regression Approach, *International Journal of Climatology* **21**, 1305-1319
- [2] Anderson, B. D. O. and Moore, J. B., 1979, *Optimal Filtering*, Englewood Cliffs, NJ: Prentice-Hall
- [3] Ansley, C. F. and Kohn, R., 1985, Estimation, Filtering and Smoothing in State Space Models With Incompletely Specified Initial Conditions, *Annals of Statistics*, **13**, 1286-1316
- [4] Barrodale, I. and Roberts, F. D. K., 1973, An Improved Algorithm for Discrete L1 Linear Approximations, *SIAM Journal of Numerical Analysis* **10**, 511-547.
- [5] Beran J., 1989, A Test of Location for Data with Slowly Decaying Serial Correlations, *Biometrika* **76**, 261-269
- [6] Beran J., 1994, *Statistics for Long-Memory Processes*, Chapman & Hall
- [7] Bhattacharya, R. N., Gupta, V. K. and Waymire, E., 1983, The Hurst Effect under Trends, *Journal of Applied Probability* **20**, 649-662
- [8] Bloomfield, P., 1976, *Fourier Analysis of Time Series: An Introduction* Wiley: Interscience
- [9] Box, G. E. P., Jenkins, G. M., 1970, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day
- [10] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., 1994, *Time Series Analysis: Forecasting and Control*, Englewood Cliffs, NJ: Prentice-Hall

- [11] Brockwell, P. J. and Davis, R. A., 1987, *Time Series: Theory and Models*, Springer, New York
- [12] Brockwell, P. J. and Davis, R. A., 1996, *Introduction to Time Series and Forecasting*, Springer, New York
- [13] Brogan, W. L., 1991, *Modern Control Theory*, Prentice Hall
- [14] Caballero, R., Jewson, S. and Brix, A., 2001, Long Memory in Surface Air Temperature: Detection, Modelling, and Application to Weather Derivative Valuation, *Climate Research* **21**, 127-140
- [15] Caines, P. E., and Mayne, D. Q., 1970, On the Discrete Time Matrix Riccati Equation of Optimal Control, *International Journal of Control* **12**, 785-794
- [16] Chan, S.W., Goodwin, G.C. and Sin, K.S., 1984, Convergence Properties of the Riccati Difference Equation in Optimal Filtering of Non-stabilizable Systems, *IEEE Transactions on Automatic Control* **29**, 110-118
- [17] DeGaetano, A. T. and Eggleston, K. L. and Knapp, W. W., 1995, A Method to Estimate Missing Daily Maximum and Minimum Temperature Observations, *Journal of Applied Meteorology* **34**, 363-387
- [18] De Jong, P., 1988, A Cross-Validation Filter for Time Series Models, *Biometrika* **75**, 594-600
- [19] De Jong, P., 1989, Smoothing and Interpolation with the State-Space Model, *Journal of the American Statistical Association* **84**, 1085-1088
- [20] De Jong, P., 1991, The Diffuse Kalman Filter, *Annals of Statistics* **2**, 1073-1083
- [21] De Jong, P., 1991b, Stable Algorithms for the State Space Model, *Journal of Time Series Analysis* **12**, 143-157
- [22] De Jong, P. and Penzer, J., 1998, Diagnosing Shocks in Time Series, *Journal of the American Statistical Association* **93**, 796-806
- [23] De Jong, P. and Penzer, J., 2004, The ARMA Model in State Space form, *Statistics & Probability Letters*, **70**, Issue 1, 119-125

- [24] Doornik, J. A., 2002, , *Object-Oriented Matrix Programming Using Ox*, 3rd ed., London: Timberlake Consultants Press and Oxford: [www.doornik.com](http://www.doornik.com)
- [25] Doornik, J. A. and Ooms, M., 2004, Inference and Forecasting for ARFIMA Models with an Application to US and UK Inflation, *Studies in Nonlinear Dynamics & Econometrics* **8**, Issue 2, Article 14
- [26] Duncan, D. B. and Horn, S. D., 1972, Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis, *Journal of the American Statistical Association* **67**, 815-821.
- [27] Dunis, C. L. and Karalis, V., 2003, Weather Derivatives Pricing and Filling Analysis for Missing Temperature Data, *Derivatives Use, Trading and Regulation* **9/1**, 61-83
- [28] Durbin, J. and Koopman, S. J., 2001, *Time Series Analysis by State Space Methods*, Oxford University Press
- [29] Edelsbrunner, H., 1987, *Algorithms in Combinatorial Geometry* Springer-Verlag
- [30] Eischeid, J. K. and Baker, C. B. and Karl, T. R. and Diaz, H. F., 1995, The Quality Control of Long-Term Climatological Data using Objective Data Analysis *Journal of Applied Meteorology* **34**, 2787-2795
- [31] Fahrmeir, L. and Tutz, G., 1994, *Multivariate Statistical Modelling Based on Generalized Linear Models*, New York:Springer-Verlag.
- [32] Hamilton, J. D., 1994, *Time Series Analysis*, Princeton University Press
- [33] Harvey, A. C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press
- [34] Harvey, A. C., 1993, *Time Series Models*, Harvester Wheatsheaf
- [35] Hosking, J., 1981, Fractional Differencing, *Biometrika* **68**, 165-176
- [36] Indyk, P. and Motwani, R., 1999, Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality *Proceedings of the Thirtieth Annual ACM Symposium On theory of Computing*, 604 - 613

- [37] Kalman, R. E., 1960, A New Approach to Linear Filtering and Prediction Problems, *Transactions of the ASME – Journal of Basic Engineering* **82D**, 35-45
- [38] Kemp W. P., Burnell, D. G. and Everson, D. O. and Thompson, A. J., 1983, Estimating Missing Daily Maximum and Minimum Temperatures, *Journal of Climate and Applied Meteorology* **22**, 1744-1763
- [39] Kleinberg M. K., 1997, Two Algorithms for Nearest-Neighbor Search in High Dimensions, *Proceedings of the Twenty-Ninth Annual ACM Symposium On theory of Computing*
- [40] Kohn, R. and Ansley, C. F., 1989, A Fast Algorithm for Signal Extraction, Influence and Cross-Validation in State-Space Models, *Biometrika* **76**, 65-79
- [41] Koopman, S. J., 1993, Disturbance Smoother for State Space Models, *Biometrika* **80**, 117-126
- [42] Koopman, S. J. and Durbin, J., 2000, Fast Filtering and Smoothing for Multivariate State Space Models, *Journal of Time Series Analysis* **21**, 281-296
- [43] Lehmann, E. L. and Casella, G., 1990, *Theory of Point Estimation*, Springer
- [44] Montanari, A., Rosso, R. and Taqqu, M. S., 1997, Fractionally Differenced Arima Models Applied to Hydrologic Time Series: Identification, Estimation and Simulation, *Water Resources Research* **33**, 1035-1044
- [45] Park, C., Hernandez Campos, F., Le, L., Marron, J. S., Park, J., Pipiras, V., Smith, F. D., Smith, R. L., Trovero, M. and Zhu, Z., 2004 Long-Range Dependence Analysis of Internet Traffic, *Submitted to Statistical Science*
- [46] Pearlman, G. K., 1980, An Algorithm for the Exact Likelihood of a High-Order Autoregressive-Moving Average Process, *Biometrika* **67** (1), 232-233
- [47] Penzer, J., 2007, State Space Models of Time Series with Patches of Unusual Observations, *Journal of Time Series Analysis* **28** (5), 629-645

- [48] Proietti, T., 2003, Leave- $k$ -out Diagnostics in State-Space Models, *Journal of Time Series Analysis* **24**, 221-236
- [49] Taqqu, M. S, Teverovsky, V. and Willinger, W., 1995, Estimators for Long-Range Dependence: An Empirical Study, *Fractals* **3**, 785-798
- [50] Teverovsky, V. and Taqqu, M. S., 1997, Testing for Long-Range Dependence in the Presence of Shifting Means or a Slowly Declining Trend Using a Variance Type Estimator, *Journal of Time Series Analysis* **18**, 279-304
- [51] Zellner, A., 1962, An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests for Aggregation Bias, *Journal of the American Statistical Association* **57**, 348-368